(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 3 951 433 A1**

(12) **EUROPEAN PATENT APPLICATION**
published in accordance with Art. 153(4) EPC

(43) Date of publication:
09.02.2022 Bulletin 2022/06

(21) Application number: **20784538.9**

(22) Date of filing: **25.03.2020**

(51) International Patent Classification (IPC):
G01T 1/161 *(2006.01)* ***G06T 7/00*** *(2017.01)*
A61B 6/03 *(2006.01)* ***A61B 5/055*** *(2006.01)*

(52) Cooperative Patent Classification (CPC):
A61B 5/055; A61B 6/03; G01T 1/161; G06T 7/00

(86) International application number:
PCT/JP2020/013334

(87) International publication number:
WO 2020/203552 (08.10.2020 Gazette 2020/41)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(30) Priority: **29.03.2019 JP 2019066982**

(71) Applicant: **FUJIFILM Corporation**
Tokyo 106-8620 (JP)

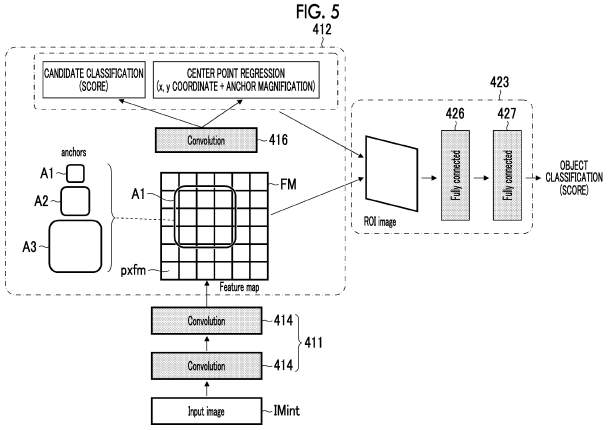
(72) Inventors:
- **KITAMURA, Yoshiro**
Tokyo 107-0052 (JP)
- **ICHINOSE, Akimichi**
Tokyo 107-0052 (JP)

(74) Representative: **Klunker IP**
Patentanwälte PartG mbB
Destouchesstraße 68
80796 München (DE)

(54) LINE STRUCTURE EXTRACTION DEVICE, METHOD AND PROGRAM, AND LEARNED MODEL

(57) Provided are a linear structure extraction device, a method, a program, and a learned model which can detect a linear structure in an image. A linear structure extraction device according to an embodiment of the present disclosure includes a learning model that is learned to receive an input of the image and output, as a prediction result, one or more element points which constitute the linear structure from the image, in which the learning model includes a first processing module that receives the image and generates a feature map representing a feature amount of the image by convolution processing, and a second processing module that calculates a shift amount from a unit center point to the element point of the linear structure closest to the unit center point, for each unit obtained by dividing the feature map into a plurality of the units including regions having a predetermined size in a grid pattern.

FIG. 5



EP 3 951 433 A1

Description

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The present invention relates to a linear structure extraction device, a method, a program, and a learned model, and particularly relates to an image processing technique and a machine learning technique of detecting a linear target object in an image.

2. Description of the Related Art

[0002] As an algorithm for detecting an object by using deep learning, in US9858496B and Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015., a method called Faster region-based convolutional neural networks (R-CNN) is proposed. In Gahayun Suh, Young-Jin Cha "Deep faster R-CNN-based automated detection and localization of multiple types of damage" Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018, a method for automatically detecting a deteriorated portion of a structure, such as rust on iron, peeling, corrosion of a bolt, and a crack in concrete, from images of a bridge and a building by using Faster R-CNN is proposed.

SUMMARY OF THE INVENTION

[0003] The crack in the concrete is a form of an object that has a linear structure. Another example of the object that has the linear structure is a tubular structure, such as a blood vessel or bronchus in a medical image. In US9858496B and Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015., an application method for detecting the linear structure from the image is not disclosed. For detecting the linear structure, it is conceivable to detect a linear object from the image by an image segmentation method. However, for the machine learning for realizing a task of image segmentation, a large number of images with correct labels for a pixel unit is required, and thus creation of such correct images is difficult.

[0004] In Gahayun Suh, Young-Jin Cha "Deep faster R-CNN-based automated detection and localization of multiple types of damage" Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018, the Faster R-CNN algorithm is applied as it is to the crack in concrete, and a bounding box including the crack is continuously detected in the image. In this case, a detection result is a rectangular region indicated by the bounding box, and it is difficult to perform reconstruction processing of specifying a center line or the like representing a region of the linear object from the detection result output as such a group of the rectangular regions.

[0005] The present invention has been made in view of such circumstances, and is to provide a linear structure extraction device, a method, a program, and a learned model which can detect a linear structure in an image.

[0006] An aspect of the present disclosure relates to a linear structure extraction device that extracts element points which constitute a linear structure from an image, the device comprising a learning model that is learned to receive an input of the image and output, as a prediction result, one or more element points which constitute the linear structure from the image, in which the learning model includes a first processing module that receives the image and generates a feature map representing a feature amount of the image by convolution processing, and a second processing module that calculates a shift amount from a unit center point to the element point of the linear structure closest to the unit center point, for each unit obtained by dividing the feature map into a plurality of the units including regions having a predetermined size in a grid pattern.

[0007] In the linear structure extraction device according to another aspect of the present disclosure, the second processing module may dispose anchors, which are one or more reference shape regions having predetermined shape and size, in each of the units, and may perform, for each unit, the convolution processing by using the feature amount at a position of the unit to calculate the shift amount for moving an anchor center point of the anchor to a closest point, which is the element point of the linear structure closest to the anchor center point, and a score for determining whether or not the linear structure is present in the anchor.

[0008] In the linear structure extraction device according to still another aspect of the present disclosure, the reference shape region may be a rectangular region in a case in which the image is a two-dimensional image, and may be a rectangular parallelepiped region in a case in which the image is a three-dimensional image.

[0009] In the linear structure extraction device according to still another aspect of the present disclosure, the linear structure may be a representative line of a region having a thickness in the image, and a plurality of the anchors having different sizes may be used in accordance with the thickness of the region having the thickness.

[0010] In the linear structure extraction device according to still another aspect of the present disclosure, the linear

structure may be a representative line of a region having a thickness in the image, and the second processing module may be learned to change the size of the anchor in accordance with the thickness of the region having the thickness, which is a target.

[0011] In the linear structure extraction device according to still another aspect of the present disclosure, the linear structure may be a representative line of a region having a thickness in the image, and the second processing module may be learned to calculate, for each anchor, a deformation magnification of the anchor in a direction of at least one side of the anchor in accordance with the thickness of the region having the thickness around the closest point.

[0012] In the linear structure extraction device according to still another aspect of the present disclosure, the region having the thickness may be a tubular structure, and the representative line may be a center line along a route of the tubular structure.

[0013] In the linear structure extraction device according to still another aspect of the present disclosure, each of the first processing module and the second processing module may be configured by a neural network, the first processing module may be configured by a convolutional neural network including a plurality of convolutional layers, and the second processing module may include a convolutional layer different from that of the first processing module and be configured by a region proposal network that predicts a candidate region including the linear structure from the feature map.

[0014] The linear structure extraction device according to still another aspect of the present disclosure may further comprise a third processing module that is learned to perform classification of each point with respect to the element points of the linear structure predicted by the second processing module.

[0015] In the linear structure extraction device according to still another aspect of the present disclosure, a class classified by the third processing module may include at least one of a root, a branch, a tip, or a point on a twig in a tree structure of graph theory.

[0016] In the linear structure extraction device according to still another aspect of the present disclosure, the linear structure may be a center line along a route of a blood vessel, and a class classified by the third processing module may include a specific anatomical name in a vascular structure.

[0017] In the linear structure extraction device according to still another aspect of the present disclosure, the linear structure may be a center line along a route of a trachea, and a class classified by the third processing module may include a specific anatomical name in a tracheal structure.

[0018] In the linear structure extraction device according to still another aspect of the present disclosure, the third processing module may be configured by a neural network, and the third processing module may include a region-of-interest pooling layer that cuts out a local image of the anchor including the element point predicted by the second processing module from the feature map and deforms the local image to a fixed size, and at least one of a convolutional layer or a fully bonded layer to which the local image deformed into the fixed size is input.

[0019] Still another aspect of the present disclosure relates to a linear structure extraction method that extracts element points which constitute a linear structure from an image, the method comprising by using a learning model that is learned to receive an input of the image and output, as a prediction result, one or more element points which constitute the linear structure from the image, receiving the input of the image to the learning model, generating a feature map representing a feature amount of the image by performing convolution processing on the input image by using a first processing module, and dividing the feature map into a plurality of units including regions having a predetermined size in a grid pattern and calculating, for each unit, a shift amount from a unit center point to the element point of the linear structure closest to the unit center point by using a second processing module.

[0020] The linear structure extraction method according to still another aspect of the present disclosure may further comprise deleting, from a point cloud of a plurality of the element points predicted by the plurality of units, a part of excess element points close to each other by less than a first interval, which is half of the size of the unit as a standard, and selecting and leaving the element points at a degree of the first interval.

[0021] In the linear structure extraction method according to still another aspect of the present disclosure, the linear structure may be a representative line of a region having a thickness in the image, and the linear structure extraction method may further comprise deleting, from a point cloud of a plurality of the element points predicted by the plurality of units, a part of excess element points close to each other by less than a second interval, which is half of the thickness as a standard, and selecting and leaving the element points at a degree of the second interval.

[0022] The linear structure extraction method according to still another aspect of the present disclosure may further comprise deleting, from a point cloud of a plurality of the element points predicted by the plurality of units, an isolation point around which other points are not present within a distance of a predetermined threshold value.

[0023] Still another aspect of the present disclosure relates to a program that causes a computer to realize a function of extracting element points which constitute a linear structure from an image, the program causing the computer to realize a function of receiving an input of the image, a function of performing convolution processing on the input image by using a first processing module and generating a feature map representing a feature amount of the image, and a function of dividing the feature map into a plurality of units including regions having a predetermined size in a grid pattern and predicting, for each unit, a shift amount from a center point of the unit to a closest element point of the linear structure

by using a second processing module.

[0024] Still another aspect of the present disclosure relates to a learned model that is learned to output, as a prediction result, one or more element points which constitute a linear structure from an input image, the model comprising a first processing module that receives the image and generates a feature map representing a feature amount of the image by convolution processing, and a second processing module that calculates a shift amount from a unit center point to the element point of the linear structure closest to the unit center point, for each unit obtained by dividing the feature map into a plurality of the units including regions having a predetermined size in a grid pattern.

[0025] In the learned model according to still another aspect of the present disclosure, parameters of networks which configure the first processing module and the second processing module may be decided by executing the machine learning by using a plurality of pieces of learning data in which a training image and positional information of the linear structure included in the training image are combined.

[0026] In the learned model according to still another aspect of the present disclosure, the linear structure may be a representative line of a region having a thickness in the image, and the learning data further may include thickness information of the region having the thickness included in the training image.

[0027] Still another aspect of the present disclosure relates to a linear structure extraction device comprising a processor, and a non-temporary computer-readable medium that stores a command for causing the processor to execute processing of extracting element points which constitute a linear structure from an image, in which the processor executes the command to perform processing including receiving the input of the image, generating a feature map representing a feature amount of the image by performing convolution processing on the input image by using a first processing module, and dividing the feature map into a plurality of units including regions having a predetermined size in a grid pattern and calculating, for each unit, a shift amount from a unit center point to the element point of the linear structure closest to the unit center point by using a second processing module.

[0028] According to the present invention, it is possible to predict the element points of the linear structure included in the image by using the learning model, and can detect the linear structure by the point cloud of the element points. According to the present invention, it is possible to easily reconstruct the linear structure from the point cloud of the predicted element points. For learning the learning model, a coordinate of the point on the line of the linear structure of a correct to the training image need only be used, and it is relatively easy to create such correct data.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029]

Fig. 1 is an example of a volume rendering (VR) image obtained by a cardiac CT examination.
Fig. 2 is a schematic diagram of a blood vessel route represented by using a node and an edge.
Fig. 3 is an example of a curved planer reconstruction (CPR) image of a coronary artery.
Fig. 4 is a configuration diagram showing an outline of a Faster R-CNN, which is applied to an embodiment of the present invention.
Fig. 5 is an explanatory diagram schematically showing contents of processing by a linear structure extraction device according to the embodiment of the present invention.
Fig. 6 is a diagram schematically showing an example of a positional relationship between each pixel of a feature map processed by a region proposal network (RPN) and a blood vessel center line.
Fig. 7 is an enlarged diagram of a unit in a vicinity of a center line CLbv.
Fig. 8 is an explanatory diagram of an anchor.
Fig. 9 is a diagram showing an example in which three types of the anchors having different sizes are used.
Fig. 10 is a conceptual diagram showing an output example of the RPN.
Fig. 11 is an explanatory diagram of an isolation point.
Fig. 12 shows an example of a point cloud provided with labels of constituting elements of a tree structure.
Fig. 13 is an explanatory diagram schematically showing a network structure of a learning model implemented on the linear structure extraction device, and a flow of processing.
Fig. 14 is a flowchart showing an example of the contents of the processing by the linear structure extraction device.
Fig. 15 is a flowchart showing an example of the contents of the processing applied to step S54 of Fig. 14.
Fig. 16 is a conceptual diagram of the learning data.
Fig. 17 is a functional block diagram showing a configuration example of a learning device which performs machine learning.
Fig. 18 is a flowchart showing an example of a learning method of the learning model in the linear structure extraction device according to the present embodiment.
Fig. 19 is a block diagram showing an example of a hardware configuration of a computer.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0030] A preferred embodiment of the present invention will be described below in detail with reference to the accompanying drawings.

«Example of Detection of Tubular Structure In Medical Image»

[0031] An example of a case in which a tubular structure is detected from a medical image will be described as an embodiment of the present invention. In recent years, with the progress of a medical device such as a multi-slice computed tomography (CT) device, a high-quality three-dimensional image is used for an image diagnosis. However, the three-dimensional image is configured by a large number of slice images and has a large amount of information, so that in some cases, it may take time for a doctor to find and diagnose a desired observation site.

[0032] Therefore, by extracting an organ of interest from the three-dimensional image and performing a display such as maximum intensity projection (MIP), volume rendering (VR), or curved planer reconstruction (CPR), the visibility of the entire organ and/or a lesion is enhanced, and the efficiency of the diagnosis is improved. For example, in a case in which analysis with respect to a cardiac CT image, particularly analysis with respect to a coronary artery or analysis with respect to a cerebral blood vessel is performed, it is required to extract a blood vessel route from the image.

[0033] Fig. 1 is an example of a VR image obtained by a cardiac CT examination. An image HVR1 shown on the left side of Fig. 1 is an example of a cardiac VR image, and an image HVR2 shown on the right side is an example of the cardiac VR image in which a coronary artery route Car is superimposed and displayed.

[0034] Fig. 2 is a schematic diagram of the blood vessel route. The blood vessel route can be represented by using a point cloud of coordinate points (node Nd) that continuously trace a center line CLbv of the blood vessel and an edge Eg representing an adjacency relationship between the nodes Nd. In a case in which the center line CLbv of the blood vessel is detected from the three-dimensional image, the plaque accumulated in the blood vessel can be visualized and a stenosis rate can be measured by generating a CPR image developed along the route, so that information useful for the diagnosis can be obtained.

[0035] Fig. 3 is an example of the CPR image of the coronary artery. An example of the CPR image in a straight view mode is shown in the lower part of Fig. 3, and a graph of an average diameter for each position along a route of a blood vessel BV is displayed in the upper part of Fig. 3. In the CPR image shown in Fig. 3, a portion in which a part of the blood vessel BV is swollen in white is plaque PLQ.

«Outline of Linear Structure Extraction Device»

[0036] A linear structure extraction device according to the embodiment of the present invention is applied to processing of extracting a center line of a tubular structure applied to support the image diagnosis as described with reference to Figs. 1 to 3. Here, as a specific application example, it is assumed that a structure of a portal vein and a vein of a liver is detected from an abdomen three-dimensional CT image. The blood vessel is an example of a "tubular structure" in the present disclosure, and the center line of the blood vessels is an example of a "linear structure" in the present disclosure. The linear structure extraction device according to the present embodiment predicts a collection of points which constitute the center line of the blood vessels from an input image, that is, a plurality of the points on the center line, and performs labeling of the classification of each point. A point cloud on the center line may be paraphrased as a "point sequence".

[0037] The linear structure extraction device according to the present embodiment improves a framework of a Faster R-CNN, which is an object detection algorithm, and performs processing of predicting the points on the line, which constitute the linear structure from the image. That is, the object, which is a detection target, in the present embodiment is the center line of the blood vessels, and an output as a prediction result is positional information of the points that are the elements which constitute the center line, that is, the points on the center line. The points that are the elements which constitute the linear structure are referred to as "element points of the linear structure". Hereinafter, the element points of the center line are referred to as the "points on the center line".

[0038] In the present specification, the term "object" is not limited to a physically existing "substance", and includes, for example, the concept of a representative line of a region, such as a crack and a gap, or a region having a size, the linear structure, and the element points of the linear structure. The center line of the blood vessels is an example of the representative line of the tubular structure having the thickness. Since the description of the processing by using the three-dimensional image is complicated, hereinafter, the description will be made by replacing with a case in which the input image is a two-dimensional image in order to facilitate understanding.

[Outline of Faster R-CNN]

[0039] Fig. 4 is a configuration diagram showing an outline of the Faster R-CNN, which is applied to the embodiment of the present invention. A Faster R-CNN 40 is configured to include a first neural network 41 that finds a region in which the object is likely to be present from the image of an input image IMipt, and a second neural network 42 that performs classification processing of specifying what the object in a candidate region RP, for each candidate region RP obtained as an output of the first neural network 41.

[0040] The first neural network 41 is configured to include a deep convolutional neural network (DCNN) 411 and a region proposal network (RPN) 412. The DCNN 441 is a neural network that extracts a feature amount of the input image IMipt. A size of a filter and the number of channels used for convolution of the DCNN 441 can be appropriately designed. For example, the filter may be a 3×3 filter, and the number of channels in a hidden layer may be 256, 512, or the like.

[0041] In a case in which the input image IMipt is input to the DCNN 411, a feature map FM is output from the DCNN 411. The feature map FM is a convolution feature map obtained by multi-layer convolution operation. The DCNN 411 may include a pooling layer, or may reduce the size of the feature map FM by setting the stride of the convolution filter to 2 or the like without providing the pooling layer. The feature map FM output from DCNN 411 is input to the RPN 412.

[0042] The RPN 412 uses the feature map FM output from the DCNN 411 as an input, and predicts the candidate region RP, which seems to be the object, from the feature map FM. The RPN 412 is configured to include a convolutional layer, and generates a bounding box (Bbox) including a thing, which seems to be the object, from the image. A list of the candidate regions RP predicted by the RPN 412 is transmitted to the second neural network 42. That is, the RPN 412 lists a plurality of the candidate regions RP from the image and transmits the listed candidate regions RP to an R-CNN 423.

[0043] The second neural network 42 is configured by the region-based convolutional neural network (R-CNN) 423. The R-CNN 423 performs the classification with respect to each candidate region RP obtained as the output of the RPN 412. In addition to the task of the classification, the R-CNN 423 may output the bounding box that represents a rectangle surrounding the object. Note that the term rectangle is not limited to a rectangle having a long side and a short side, but also includes a square.

[0044] The R-CNN 423 is connected to the DCNN 411, and the feature map FM, which is output from the DCNN 411, is input to the R-CNN 423. Further, data of the candidate region RP predicted by the RPN 412 is input to the R-CNN 423. The R-CNN 423 projects the candidate region RP generated by RPN 412 onto the feature map FM, cuts out a region of interest (ROI), which is an operation target, performs the classification of the object for each ROI, and decides a label. The R-CNN 423 performs the classification with respect to each candidate region RP obtained as the output of the RPN 412.

[0045] Further, the R-CNN 412 may output the bounding box that surrounds the detected object. In the task of the object detection by using, as a target, a subject captured in a general photographic image disclosed in US9858496B and Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015., an output of an object label and an output of the bounding box that represents a bounding rectangle of the object are obtained from the R-CNN 423.

[0046] In a case of the present embodiment, it is considered that even in a case in which the rectangle that surrounds the "points on the center line", which is the element points of the linear structure is output, it is poor in practice use, and thus the output of the bounding box may be omitted. Alternatively, if necessary, the bounding box including the points on the center line and a region range at a degree of a thickness of the blood vessel around the center line may be output.

[Contents of Processing in Linear Structure Extraction Device]

[0047] Fig. 5 is an explanatory diagram schematically showing contents of processing in a linear structure extraction device 50 according to the embodiment of the present invention. In Fig. 5, the elements common to those in Fig. 4 are designated by the same reference numerals. The linear structure extraction device 50 can be realized by a calculation system (computing system) configured by using one or a plurality of computers. The linear structure extraction device 50 comprises the DCNN 411 as a first processing module, the RPN 412 as a second processing module, and the R-CNN 423 as a third processing module. The term "module" includes the concept of a program module.

[0048] The DCNN 411 receives the input image IMipt, performs convolution processing by a plurality of convolutional layers 414, and generates the feature map FM. A leading input layer of the DCNN 411 has a role of an image reception unit that receives the input image IMipt. 6×6 grid shown in Fig. 5 represents a part of the feature map FM, and one compartment of a square of the grid corresponds to a pixel pxfm of the feature map FM. One pixel pxfm of the feature map FM has information on the feature amount calculated from a wide-range image region in the input image IMipt.

[0049] In a case in which the pixel of the input image IMipt is expressed as a pixel px, the DCNN 411 is configured in which one pixel pxfm of the feature map FM has the feature amount calculated from a pixel region having an $S \times S$ pixel size in a grid array of the pixels px of the input image IMipt. S is a value in accordance with a reduction rate of the image

by the DCNN 411. That is, it is understood that each pixel pxfm of the feature map FM corresponds to the pixel region having the S × S size at the corresponding position in the input image IMipt.

[0050] Stated another way, each region having the S × S size in a case in which the input image IMipt is divided into a plurality of regions having the S × S size in a grid pattern corresponds to the pixel pxfm of the feature map FM. The position of each pixel pxfm of the feature map FM can be described by projecting the position thereof onto a coordinate system that represents an image position in the input image IMipt.

[0051] The RPN 412 receives the feature map FM output from the DCNN 411, and predicts the point on the center line closest to a center point of the pixel pxfm, for each pixel pxfm of the feature map FM. The point on the center line closest to the center point of the pixel pxfm is referred to as a "closest point". Each pixel pxfm of the feature map FM is a unit of a region for predicting a shift amount to the closest point, and is referred to as a "unit". That is, the feature map FM is divided into a plurality of the units (pixels pxfm) having a predetermined size region in a grid pattern, and the RPN 412 predicts the closest point, which is a candidate for the element point of the center line, for each unit of the feature map FM.

[0052] The RPN 412 applies a plurality of reference rectangles having different aspect ratios and/or sizes to each unit of the feature map FM in order to predict the closest point for each unit of the feature map FM. The reference rectangle is referred to as an "anchor". Fig. 5 shows an example in which three types of the anchors having different sizes, an anchor A1, an anchor A2, and an anchor A3 are used. Here, the aspect ratios of all three types of the anchors are 1:1. Note that a plurality of the anchors having the same size and different aspect ratios may be used. A plurality of the anchor A1, the anchor A2, and the anchor A3 are disposed such that each of the center points thereof coincides with the center point of the unit (pixel pxfm).

[0053] The RPN 412 has a convolutional layer 416 that calculates how much each anchor is moved (shifted) and/or deformed to approach a correct rectangle, and calculates whether or not there is the object in the anchor. The correct rectangle here is a rectangle having the point on the correct center line at the center position, the rectangle having a region size in accordance with the thickness of the blood vessel. The RPN 412 comprises the convolutional layer different from the DCNN 411. For example, the RPN 412 may have a configuration providing a smaller number of convolutional layers than the DCNN 411. The RPN 412 performs the convolution processing, for each unit, by using the feature amount of the unit position.

[0054] The RPN 412 performs, through the convolution by the convolutional layer 416, an output of the shift amount of the anchor center point and a deformation amount of the anchor for allowing the anchor to approach the correct rectangle, and an output of a score representing the certainty of two-classification indicating whether or not there is the object in the anchor. That is, the RPN 412 solves, for each anchor, "a regression problem of how the anchor is moved and/or deformed to coincide with ground truth" and "an identification problem of whether or not there is the object in the anchor". Note that the deformation amount of the anchor may be, for example, the deformation magnification in each direction of an x direction and a y direction. In a case in which similar deformation that corrects only the size without changing the aspect ratio of the anchor is performed, the deformation amount of the anchor may be the deformation magnification common to the x direction and the y direction.

[0055] The score of the two-classification indicating whether or not there is the object in the anchor is called an "objectness score". On the other hand, regression result data showing the shift amount and the deformation amount of the anchor to allow the anchor to approach the correct rectangle is collectively referred to as "Bbox offset". The RPN 412 may calculate the objectness score indicating whether or not there is the object in the anchor shifted by the calculated shift amount, may calculate the objectness score indicating whether or not there is the object in the anchor which is not shifted (in the anchor disposed at the unit position), or may calculate both the objectness scores.

[0056] The R-CNN 423 generates a local image obtained by cutting out a portion corresponding to the candidate region from the feature map FM based on the candidate region of the prediction result output from the RPN 412, calculates the score of the classification of the object included in an ROI image by fully bonded layers 426 and 427 based on the ROI image, which is the cut out local image, and provides a class label based on the score. Note that the size of the local image cut out from the feature map FM may be different from the size of the candidate region.

[0057] Hereinafter, the RPN 412 and the R-CNN 423 will be described in more detail.

[Description of RPN 412]

[0058] For example, the RPN 412, which is used for extraction of the blood vessel in the liver, is trained to present a candidate point on the center line of the blood vessels, regardless of the portal vein or the vein. Each pixel of the feature map obtained from a final layer of RPN 412 is a unit that predicts the candidate point on the center line. Each unit corresponds to the pixel pxfm of the feature map FM described with reference to Fig. 5. Each unit predicts a deviation amount from the position of the closest point of the center line with respect to the center position of the unit.

[0059] In the case of the two-dimensional image, the deviation amount predicted by the RPN 412 is two real values Δx and Δy indicating the deviation amount in each of the x direction and the y direction. In the case of the three-dimensional

image, the deviation amount predicted by the RPN 412 is three real values Δx, Δy, and Δz indicating the deviation amount in each of the x direction, the y direction, and the z direction. Further, the RPN 412 simultaneously determines whether or not there is a target object in the anchor shifted in accordance with the predicted deviation amount and/or in the anchor which is not shifted. That is, the RPN 412 performs the two-classification indicating whether or not there is the target object.

[0060] Fig. 6 is a diagram schematically showing an example of a positional relationship between each pixel of the feature map processed by the RPN 412 and a blood vessel center line. 8 $\times$ 8 grids shown in Fig. 6 represent a part of the feature map FM from the DCNN 411. For example, in Fig. 6, in a case in which a position of each unit u is expressed as "u(column number, row number)" by using column numbers 0 to 7 and row numbers 0 to 7 of the grid, the upper left unit is expressed as u(0, 0), the lower right unit is expressed as u(7, 7). For example, the unit u(3, 4) predicts a deviation amount (Δx, Δy) between a center coordinate CP34 of the unit center position and a position of a closest point NP34 of the center line CLbv closest to the center coordinate CP34.

[0061] Note that a space that defines the coordinate may be an xy coordinate system that specifies the position in the input image IMipt. That is, the center coordinate and the closest point coordinate of each unit are represented by numerical values (x, y) in the xy coordinate system that specify the position in the image of the input image IMipt. Similarly for the other units, the deviation amount (Δx, Δy) from the position of the closest point of the center line CLbv is predicted with reference to the center coordinate of each unit. The center coordinate of the unit u is an example of a "unit center point" in the present disclosure. The predicted "deviation amount" is an example of a "shift amount" from the unit center point to the closest point in the present disclosure.

[0062] Fig. 7 is an enlarged diagram of a unit of the center line CLbv. Here, four of the units u are shown. The square of the fine grid displayed in the unit u schematically represents a pixel size of the input image IMipt. A closest point NP of the center line CLbv is predicted from a center coordinate CP of each unit u as shown in Fig. 7.

<Description of Anchor>

[0063] Fig. 8 is an explanatory diagram of the anchor. Fig. 8 shows a part of the feature map FM as in Fig. 6, and the grid square represents the pixel of the feature map FM, that is, the unit. Each unit imaginarily has a plurality of predefined anchors. Fig. 7 shows two types of the anchors for simplifying the description. Fig. 7 shows a first anchor 71 and a second anchor 72 disposed in the unit u(4, 4) filled with gray. The first anchor 71 is an anchor having a pixel size of 3 $\times$ 3. The second anchor 72 is an anchor having a pixel size of 7 $\times$ 7.

[0064] It is assumed that the thickness of the blood vessel is defined in accordance with the position on the center line CLbv of the blood vessel. Among the plurality of anchors disposed in each unit, only the anchor having the size which most approaches a target thickness predicts the position of the point on the center line.

[0065] A plurality of types of the anchors are prepared such that a range of the target thickness is covered. For example, in a case in which the coronary artery is a target, three-sized square anchors with sides of 3 pixels, 5 pixels, and 9 pixels are prepared. Note that in a case in which the object detection is performed with a general object as a target disclosed in Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015., the plurality of anchors having different aspect ratios are prepared, but in a case of the present embodiment, the detection target is the tubular structure having the linear structure extending in various directions, there is no particular tendency for directionality such as vertical or horizontal, and thus the aspect ratio of the anchor may be only 1:1.

[0066] In a case of Fig. 8, a vicinity of the lower left of the center line CLbv, for example, a vicinity of the unit u(1, 5) has the thickness of the blood vessel of approximately 1 pixel in radius. On the other hand, a vicinity of the unit u(4, 3) positioned near the upper right of the center line CLbv has the thickness of the blood vessel of approximately 2 pixels in radius. Therefore, in the unit disposed in the lower left portion of Fig. 8, among the plurality of anchors, the anchor having a pixel size of 3 $\times$ 3 is used for the prediction of the position of the point on the center line, and in the unit disposed in the upper right portion of Fig. 8, the anchor having a pixel size of 7 $\times$ 7 is used for the prediction of the position of the point on the center line.

[0067] Fig. 9 shows an example in which three types of the anchors having different sizes are used, and shows that the anchor used for the prediction of the point on the center line is decided in accordance with the thickness of the blood vessel. In this example, an anchor 81 having a side of 3 pixels, an anchor 82 having a side of 5 pixels, and an anchor 83 having a side of 7 pixels are prepared, and the size of the anchor to be applied is changed in accordance with the thickness of the blood vessel. Note that it is not necessary to limit the number of the anchors used for the prediction for one unit to one, and depending on the thickness of the blood vessel, the plurality of anchors may be applied to one unit to predict the point on the center line.

[0068] As the learning data used for learning, the positional information of the correct center line is provided to the training image, and information indicating what size of a region (here, the thickness of the blood vessel) each point on the correct center line represents is provided. That is, as the learning data, information (score) indicating that each point on the correct center line is to be extracted by the anchor of what size, that is, is to be extracted as a representative

point of what thickness is provided. As a result, it is possible to learn to change the size of the anchor in accordance with the thickness of the blood vessel of the region, which is a target.

[0069] Fig. 10 is a conceptual diagram showing an output example of the RPN 412. Fig. 10 shows an example of the candidate regions RP obtained by performing, for each anchor, shift of the center point coordinate to the predicted closest point NP and rectangular size correction by the predicted deformation magnification shown in Fig. 9.

<Outline of Learning Method By Using Anchor>

[0070] An example of a procedure of a learning method by using the anchor is described below.

[Step 1] The RPN 412 disposes the plurality of predefined anchors in each unit (pixel) of the feature map FM output from the DCNN 411 by an input of the training image.
[Step 2] The RPN 412 searches for the anchor having a large overlap with the correct rectangle from among the plurality of anchors.
[Step 3] A difference between a selection anchor selected in step 2 and the correct rectangle is calculated. This difference may be, specifically, the shift amounts Δx and Δy of the anchor center coordinate and the deformation magnification for changing the size of the anchor.
[Step 4] The network is learned such that the objectness score of the selection anchor is "1" and a correction amount of the bounding box (Bbox offset) is the difference calculated in step 3.

<Outline of Inference Method By Using Anchor>

[0071] An example of an inference (prediction) method by using the anchor is described below.

[Step 101] The learned RPN 412 disposes the plurality of predefined anchors in each unit of the feature map FM output from the DCNN 411 by an input of an unknown image, which is an inference target.
[Step 102] The RPN 412 calculates the Bbox offset of each anchor and the objectness score.
[Step 103] For the anchor having a high objectness score, the anchor is moved and deformed based on the Bbox offset of the anchor.

<Suppression of Overlapping Candidate Regions: Non-Maximum Suppression (NMS) Processing>

[0072] In some cases, the point cloud of the points on the center line predicted by each unit is excessively large. As disclosed in US9858496B and Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015., in the Faster R-CNN, NMS processing of selecting and leaving only an important candidate is inserted between the RPN and the R-CNN. The NMS processing is processing of leaving one rectangle among a plurality of rectangles indicating the same object and suppressing an output from another rectangle.

[0073] In cases of US9858496B and Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015., the RPN calculates an intersection over union (IoU) value between the generated candidate regions, and in a case in which the IoU value is larger than a predetermined threshold value, it is considered that the overlap between the regions is large, and one region is deleted (suppressed). On the contrary, in a case in which the IoU value is small, the overlap between the regions is small, and thus both candidate regions are left as it is. A mechanism of reducing the number of excessively overlapping candidate regions is proposed by such an algorithm.

[0074] In a case of a problem of detection of the point cloud on the center line, which is a target in the present embodiment, it is sufficient that the "points on the center line" are detected at intervals of a degree of half of the thickness of the blood vessel. Therefore, in the present embodiment, in addition to the NMS processing described above or instead of the NMS processing, the IoU value described above is not calculated, and processing of thinning out the candidate regions is performed at intervals of a degree of half of the thickness of the blood vessel. Note that in a case in which the information on the thickness of the blood vessel is not provided in advance as teaching data in a case of learning, sampling need only be performed at a degree of pixel intervals of the unit.

<Removal of Isolation Point>

[0075] Fig. 11 shows an example of the candidate point predicted by the RPN 412, and is an explanatory diagram in a case in which the isolation point is included in the candidate point. Since the linear structure such as the center line of the blood vessels is represented by the array of continuous points (point sequence), as shown in Fig. 11, in a case

in which the point on the center line predicted from each unit is present at a large distance from the other point sequences and is isolated, there is a high possibility that an isolation point ISP is a falsely predicted result (false detection). Therefore, in the RPN 412, in a case in which the predicted candidate point on the center line is isolated, a predetermined threshold value that can be determined as the false detection is set, and in a case of a point around which other candidate points are not present in a distance within the threshold value, the isolation point ISP thereof is deleted (removed) from the prediction result.

[Description of R-CNN 423]

[0076] The R-CNN 423 performs the classification by using, as an input, the image obtained by standardizing the feature map in the anchor predicted by the RPN 412. In a case in which the detection target is a tree structure of graph theory, such as a vascular structure handled in the present embodiment, the R-CNN 423 performs the classification into any one of four labels as a "root", a "point on a twig", a "branch point", or a "peripheral point (tip)" as constituting elements of the tree structure.

[0077] Fig. 12 shows an example of the point cloud provided with the labels of the constituting elements of the tree structure. In a case in which a characteristic (classification) of each point is known in advance in this way, it is convenient to connect the points in the subsequent stage to reconstruct a graph structure. For example, a search for the route can be started from a position of the root, the number of twigs can be increased at the branch point, and the route connection can be terminated at the peripheral point.

[0078] The existing algorithm, such as a minimum spanning tree algorithm or a shortest route (Dijkstra) algorithm, can be used for the route connection.

<Other Examples of Classification>

[0079] There are various vascular systems in a human body, such as a liver or lungs. The vascular system of the liver includes an artery, the portal vein, and the vein. The respective vascular systems are in contact with each other or intersect each other, and it is important to separate the target in order to understand the anatomy. Therefore, a configuration may be adopted in which the R-CNN 423 performs the classification of the types of the blood vessels. In this case, an anatomical name need only be provided as an item of a class to be classified, and the data of the label, which is the correct, need only be added to the learning data.

[Case of Liver]

[0080] For the purpose of classification of the blood vessels in the liver, the classes classified by the R-CNN 423 for the candidate points (points on the predicted center line) detected by the RPN 412 are set to four classes of {the portal vein, the vein, the artery, and others}, depending on the types of the blood vessels.

[0081] Further, the liver is anatomically divided into eight areas. The eight areas are caudate lobe, a lateral dorsal area, a lateral caudal area, a medial area, an anterior rostral area, an anterior caudal area, a posterior occipital area, and a posterior caudal area. Since these areas are defined by running of vascular twigs, eight types of the vascular twigs can be classified.

[0082] For the provided correct center line, the anatomical name is provided to each twig. The label of the candidate point on the predicted center line is used as the correct label learned by the R-CNN 423.

[0083] In the human body, in addition to the liver, there is a tree structure (called "graph" in a broad sense in a case in which a loop is present), such as cerebral blood vessels, pulmonary blood vessels, bronchi, or gastrointestinal tract. A method in the present disclosure can be applied to the recognition of various anatomical structures.

[Case of Lungs]

[0084] In a case of the pulmonary vascular structure, for example, a pulmonary vein and a pulmonary artery can be classified. Alternatively, for the tracheal structure having tracheal and bronchial tree structures, a plurality of classifications can be performed by an anatomical bronchial name and/or an area name. The lungs are divided into a plurality of areas by bronchial twigs. For example, the classification can be made into classes of the trachea, the main bronchus of the right lung, the upper lobe bronchus, the apical bronchus (B1), the posterior bronchus (B2), the anterior bronchus (B3), the truncus intermedius, the middle lobe bronchus, the lateral bronchus (B4), the medial bronchus (B5), the lower lobe bronchus, the superior bronchus (B6), the medial basal bronchus (B7), the anterior basal bronchus (B8), the lateral basal bronchus (B9), the posterior basal bronchus (B10), the basal bronchus, the main bronchus of the left lung, the upper lobe bronchus, upper segment bronchus, the apical posterior bronchus (B1+2), the anterior bronchus (B3), the ramus lingularis, the ramus lingularis superior (B4), the ramus lingularis inferior (B5), the lower lobe bronchus, the superior

bronchus (B6), the medial anterior basal bronchus (B7+8), the lateral basal bronchus (B9), the posterior basal bronchus (B10), the basal bronchus, and the like.

«Example of Learning Model Used In Linear Structure Extraction Device»

[0085] Fig. 13 is an explanatory diagram schematically showing a network structure of a learning model 52 implemented on the linear structure extraction device 50, and a flow of processing. The elements corresponding to the elements described in Figs. 4 and 5 are designated by the same reference numerals in Fig. 13, and the description thereof will be omitted. The learning model 52 includes the DCNN 411, the RPN 412, and the R-CNN 423.

[0086] The convolutional layer 416 of the RPN 412 has the number of filters in accordance with the number of channels of the feature map FM output by the DCNN 411. For example, the filter size of the convolutional layer 416 may be 3×3.

[0087] The RPN 412 has two types of 1×1 convolutional layers 417 and 418 in the subsequent stage of the convolutional layer 416. An output of the convolutional layer 416 is input to each of the 1×1 convolutional layers 417 and 418. The 1×1 convolutional layer 417 includes a softmax layer that uses a softmax function as an activation function, and outputs the objectness score indicating a probability of being the object (point on the center line) at each anchor position. The 1×1 convolutional layer 418 is a regression layer that performs numerical regression to allow the anchor to approach the correct rectangle for each of the plurality of anchors. The RPN 412 is trained such that the overlap with the correct rectangle of the training data is large.

[0088] The R-CNN 423 includes an ROI pooling layer 424, fully bonded layers 426 and 427, and a softmax layer 428. The ROI pooling layer 424 pools the feature map in the region corresponding to each candidate region RP cut out from the feature map FM obtained from the DCNN 411, and deforms the pooled feature map into a standardized image having a fixed size. A partial image of the feature map deformed into the fixed size is input to the fully bonded layer 426. The softmax layer 428 is provided in the subsequent stage of the final fully bonded layer 427. The number of units in an output layer is decided in accordance with the number of classes to be classified, the object score indicating the probability of being each class is calculated, and finally the object label is specified. Note that a configuration may be adopted in which the convolutional layer is provided in place of or in addition to a part or all of the fully bonded layers 426 and 427.

[0089] The ROI pooling layer 424 shown in Fig. 13 is an example of a "region-of-interest pooling layer" in the present disclosure. The learning model 52 is an example of a "learned model" in the present disclosure.

<Linear Structure Extraction Method According To Present Embodiment>

[0090] Fig. 14 is a flowchart showing an example of the contents of the processing by the linear structure extraction device 50. The processing shown in Fig. 14 is executed by the calculation system that functions as the linear structure extraction device 50. The calculation system executes the processing of each step in accordance with a program stored on a computer-readable medium.

[0091] In step S50, the calculation system receives the image, which is a processing target.

[0092] In step S52, the calculation system generates the convolution feature map by the DCNN 411 from the input image.

[0093] In step S54, the calculation system inputs the convolution feature map, which is output from the DCNN 411, to the RPN 412, and generates the candidate region that seems to be the point on the center line by the RPN 412.

[0094] In step S56, the calculation system inputs the information on each candidate region generated by the RPN 412 and the convolution feature map generated by the DCNN 411 to the R-CNN 423, cuts out each candidate region by the R-CNN 423, and generates a classification label of the object of each candidate region.

[0095] In step S58, the calculation system stores the data of the prediction result in which the position of each point on the center line predicted by the RPN 412, the thickness of the blood vessel of each point, and the label of each point predicted by the R-CNN 423 are associated.

[0096] After step S58, the calculation system terminates the flowchart of Fig. 14.

[0097] Fig. 15 is a flowchart showing an example of the contents of the processing applied to step S54 of Fig. 14. In step S61 of Fig. 15, the calculation system generates the plurality of anchors for each unit of the convolution feature map.

[0098] In step S62, the calculation system predicts, for each anchor, the coordinate of the point (closest point) on the center line of the blood vessels, which is closest to the anchor center.

[0099] In step S63, the calculation system calculates the objectness score for the two-classification indicating whether or not the center line point is included in each anchor.

[0100] In step S64, the calculation system predicts, for the anchor having a high objectness score, the anchor magnification in accordance with the thickness of the blood vessel at the position of the predicted closest point.

[0101] In step S65, the calculation system suppresses the excess candidate region from a large number of candidate regions generated by the RPN 412 in consideration of the thickness of the blood vessel. For example, sampling is performed such that in the point cloud of a plurality of the candidate points predicted by the RPN 412, a part of excess

candidate points close to each other by less than an interval (first interval), which is a degree of half of a diameter of the blood vessel (radius) as a standard, is deleted, and the candidate points at the interval of a degree of the radius of the blood vessel are selected and left. By such thinning sampling, the point sequence of the candidate points remains at a larger interval as the thickness of the blood vessel is thicker, and the candidate points remain at a smaller interval as the thickness of the blood vessel is thinner.

[0102] Note that in a case in which the information on the thickness of the detection target object is not provided in advance, sampling is performed such that in the point cloud of a plurality of the candidate points predicted by the RPN 412, a part of excess candidate points close to each other by less than an interval (second interval), which is a degree of half of a size of the unit u as a standard, is deleted, and the candidate points at the interval of a degree of half of the size of the unit u are selected and left.

[0103] In step S66, the calculation system determines the isolation point from the candidate point predicted by the RPN 412, and deletes the candidate region of the isolation point.

[0104] In step S67, the calculation system generates the data of the prediction result in which the position of each point on the predicted center line and the thickness of the blood vessel at each point are associated, that is, a list of the Bbox of the candidate region.

[0105] After step S67, the calculation system terminates the flowchart of Fig. 15, and returns to the flowchart of Fig. 14.

«Example of Learning Method»

[0106] Hereinafter, an example of the learning method of the learning model in the linear structure extraction device 50 according to the present embodiment will be described.

[Example of Learning Data]

[0107] As the learning data used for learning, a plurality of sets of the combination of the training image, the positional information of each point on the center line of the blood vessels included in the training image, the information on the thickness of the blood vessel at each point of the center line, and a correct label of the classification for each point are used. The "learning data" is data for training used for machine learning, and is synonymous with "data for learning" or "training data".

[0108] The training image may be, for example, a CT image captured by a CT device. As the positional information of each point on the provided correct center line of the blood vessels and the information on thickness of the blood vessel, for example, the coordinate of the point on the center line of the CPR image generated from the CT image and a numerical value of the blood vessel radius can be used.

[0109] By specifying the thickness of the blood vessel (radius) of each point, for example, it is possible to automatically determine the correct rectangle of the square with one side having a length twice the radius centered on the point. In addition, the size of the anchor suitable for the prediction of the point can be determined from the provided blood vessel thickness of each point. The correct label of the classification of each point can be determined based on anatomical findings. For one training image, the correct data of the position at which it is desire to extract the anchor of each size is provided for each type (size) of the anchor. Note that depending on the thickness of the blood vessel, the correct data may be provided such that duplicate predictions are performed by using the plurality of anchors of different sizes.

[0110] Fig. 16 is a conceptual diagram of learning data LD(i). In the machine learning of the present embodiment, the training image, the coordinate of each point of the correct center line, the correct thickness of each point, and the correct label of each point are provided as the learning data LD(i). i is an index number for identifying the learning data. Note that the coordinate of each point of the correct center line may be provided as a numerical value in sub-pixel units, which is finer than pixel units of the training image. The correct rectangle can be automatically generated from the information on the correct thickness. The size of the anchor may be automatically generated from the information on the correct thickness, or may be designated by an operator.

[Configuration Example of Learning Device]

[0111] Fig. 17 is a functional block diagram showing a configuration example of a learning device 100 which performs the machine learning. The learning device 100 can be realized by the calculation system configured by using one or a plurality of computers. The calculation system which constitutes the learning device 100 may be the same system as the calculation system which constitutes the linear structure extraction device 50, a different system, or a system that shares a part of elements.

[0112] The learning device 100 is connected to a learning data storage unit 150. The learning data storage unit 150 is configured to include a storage that stores the learning data LD(i) necessary for the learning device 100 to perform the machine learning. Here, an example will be described in which the learning data storage unit 150 and the learning

device 100 are respectively configured as separate devices, but these functions may be realized by one computer, or the processing functions may be allocated and realized by two or more of a plurality of computers.

[0113] For example, the learning data storage unit 150 and the learning device 100 may be connected to each other via an electric telecommunication line (not shown). The term "connection" is not limited to wired connection, and also includes the concept of wireless connection. The electric telecommunication line may be a local area network, or may be a wide area network.

[0114] With this configuration, the generation processing of the learning data and the learning processing of the learning model can be performed without being physically and temporally restricted by each other.

[0115] The learning device 100 reads the learning data LD(i) from the learning data storage unit 150 and executes the machine learning. The learning device 100 can perform reading of the learning data LD(i) and updating of the parameters in units of mini-batch in which a plurality of pieces of the learning data LD(i) are collected.

[0116] The learning device 100 includes a data acquisition unit 102, a learning model 52, a first error calculation unit 110, a second error calculation unit 112, and an optimizer 114.

[0117] The data acquisition unit 102 is an interface that introduces the learning data LD(i). The data acquisition unit 102 may be configured by a data input terminal that introduces the learning data LD(i) from other signal processing units outside or inside the device. Further, the data acquisition unit 102 may employ a wired or wireless communication interface unit, may employ a media interface unit that performs reading and writing of a portable external storage medium such as a memory card, or may employ an appropriate combination of these aspects.

[0118] As described above, the learning model 52 includes the DCNN 411, the RPN 412, and the R-CNN 423.

[0119] The first error calculation unit 110 calculates an error between the prediction result output from the RPN 412 and the correct data for each anchor. The first error calculation unit 110 evaluates the error by using a loss function. A first error calculated by the first error calculation unit 110 is transmitted to the optimizer 114.

[0120] The second error calculation unit 112 calculates the error between the prediction result output from the R-CNN 423 and the correct label. The second error calculation unit 112 evaluates the error by using the loss function. A second error calculated by the second error calculation unit 112 is transmitted to the optimizer 114.

[0121] The optimizer 114 performs processing of updating the parameters of the learning model 52 from calculation results of the first error calculation unit 110 and the second error calculation unit 112. The optimizer 114 performs updating of the parameters based on the algorithm such as a back propagation method. The network parameters include a filter coefficient (weight of connection between nodes) of filters used to processing each layer, node bias, and the like.

[0122] The optimizer 114 calculates an update amount of the parameter of a first subnetwork 410 obtained by connecting the DCNN 411 and the RPN 412 by using the calculation result of the first error calculation unit 110, and performs the parameter updating processing of updating the parameters of the network of at least the RPN 412 of the DCNN 411 and the RPN 412 in accordance with the calculated update amount of the parameter. Preferably, the parameters of the respective networks of the DCNN 411 and the RPN 412 are updated.

[0123] Further, the optimizer 114 calculates the update amount of the parameter of a second subnetwork 420 obtained by connecting the DCNN 411 and the R-CNN 423 by using the calculation result of the second error calculation unit 112, and updates the parameter of the network of each of the DCNN 411 and the R-CNN 423 in accordance with the calculated update amount of the parameter.

[0124] Further, the learning device 100 further performs learning for the model of the first subnetwork 410 in a state in which the parameter of the DCNN 411 finely tuned by the training of the second subnetwork 420 are fixed, and updates the parameter of the RPN 412. By repeatedly executing such a learning process, the parameters of the learning model 52 can be optimized. As described above, the learned learning model 52 can be obtained.

[Example of Learning Method By Using Learning Device 100]

[0125] Fig. 18 is a flowchart showing an example of the learning method of the learning model 52 in the linear structure extraction device 50 according to the present embodiment. The processing shown in Fig. 18 is executed by the calculation system configured by using one or a plurality of computers functioning as the learning device 100. The calculation system executes the processing of each step in accordance with a program stored on a computer-readable medium. The calculation system used for the machine learning may be the same system as the calculation system which constitutes the linear structure extraction device 50, a different system, or a system that shares a part of elements.

[0126] In step S202 of Fig. 18, the learning device 100 performs initial setting of the learning model 52. Here, the initial setting of the learning model 52 having the network structure shown in Fig. 13 is performed. The parameters of each network of the DCNN 411, the RPN 412, and the R-CNN 423 are set to initial values. A part of the parameters may be learned parameters obtained by learning in advance.

[0127] In step S204 of Fig. 18, the learning device 100 trains the model of the first subnetwork 410 obtained by connecting the DCNN 411 and the RPN 412. In step S204, the parameters of the networks of the DCNN 411 and the RPN 412 are updated. Note that the learning device 100 can acquire the learning data in units of mini-batch including

the plurality of pieces of learning data LD(i), and the optimizer 114 can perform the parameter updating processing in units of mini-batch.

[0128] Then, in step S206, the learning device 100 generates the candidate region from the training image by using the trained first subnetwork 410.

[0129] In step S208, the learning device 100 inputs the candidate region generated by the trained first subnetwork 410 to the R-CNN 423 and trains the model of the second subnetwork 420 in which the DCNN 411 and the R-CNN 423 are connected. In step S208, the parameters of the networks of the DCNN 411 and the R-CNN 423 are updated.

[0130] In step S210, the learning device 100 retrains the RPN 412 of the trained first subnetwork 410 by using the DCNN 411 of the trained second subnetwork 420.

[0131] After step S210, the learning device 100 may return to step S206 to repeat training, or may terminate the flowchart of Fig. 18 based on a predetermined learning termination condition.

[0132] The learning termination condition may be determined based on the error value or may be determined based on the number of updates of parameter. As the method based on the error value, for example, the learning termination condition may be that the error converges within a specified range. As the method based on the number of updates, for example, the learning termination condition may be that the number of updates reaches the specified number of times.

«Application to Three-Dimensional Image»

[0133] Although the description has been made so far with the two-dimensional image as an example, the matters described about the two-dimensional image can be extended and applied to the processing of the three-dimensional image. The replacement in a case of expanding from two dimensions to three dimensions is, for example, as follows.

[0134] The "pixel" can be paraphrased as a "voxel". The "rectangle" can be paraphrased as a "rectangular parallelepiped". A "cube" can be understood as a kind of a "rectangular parallelepiped". The two-dimensional xy coordinate can be paraphrased as a three-dimensional xyz coordinate. The "aspect ratio" of the rectangle can be paraphrased as a "ratio of three sides" of the rectangular parallelepiped. The anchor can be understood as a reference shape region having predetermined shape and size, and in a case of the three-dimensional image, the rectangular parallelepiped having a three-dimensional shape is used. That is, the reference shape region of the anchor for the two-dimensional image is a rectangular region, whereas the reference shape region of the anchor for the three-dimensional image is a rectangular parallelepiped region.

«Example of Hardware Configuration of Computer»

[0135] Fig. 19 is a block diagram showing an example of a hardware configuration of the computer. A computer 800 may be a personal computer, a workstation, or a server computer. The computer 800 can be used as a device having a part or all of any one of the linear structure extraction device 50, the learning device 100, or the learning data storage unit 150, which have been described, or a device having a plurality of functions thereof.

[0136] The computer 800 comprises a central processing unit (CPU) 802, a random access memory (RAM) 804, a read only memory (ROM) 806, a graphics processing unit (GPU) 808, a storage 810, a communication unit 812, an input device 814, a display device 816, and a bus 818. Note that the graphics processing unit (GPU) 808 need only be provided, if necessary.

[0137] The CPU 802 reads out various programs stored in the ROM 806, the storage 810, or the like, and executes various pieces of processing. The RAM 804 is used as a work region of the CPU 802. Further, the RAM 804 is used as a storage unit that temporarily stores the read program and various pieces of data.

[0138] The storage 810 is configured to include, for example, a hard disk apparatus, an optical disk, a magneto-optical disk, or a semiconductor memory, or a storage device configured by using an appropriate combination thereof. The storage 810 stores various programs, pieces of data, or the like required for linear structure extraction processing and/or learning processing. The program stored in the storage 810 is loaded into the RAM 804, and the CPU 802 executes the program, so that the computer 800 functions as means for performing various pieces of processing specified by the program.

[0139] The communication unit 812 is an interface that performs communication processing with an external device by wire or wirelessly, and performs exchange of the information with the external device. The communication unit 812 can have the role of the image reception unit that receives the input of the image.

[0140] The input device 814 is an input interface that receives various operation inputs with respect to the computer 800. For example, the input device 814 may be a keyboard, a mouse, a touch panel, other pointing devices, a voice input device, or an appropriate combination thereof.

[0141] The display device 816 is an output interface on which various types of information are displayed. For example, the display device 816 may be a liquid crystal display, an organic electro-luminescence (OEL) display, a projector, or an appropriate combination thereof.

<<About Program Causing Computer to Operate>>

[0142] The program that causes the computer to realize a part or all of at least one processing function of the linear structure extraction function or the learning function described in the embodiment can be recorded on the computer-readable medium which is the tangible non-temporary information storage medium such as the optical disk, the magnetic disk, the semiconductor memory, or other objects, and the program can be provided through the information storage medium.

[0143] Further, instead of the aspect in which the program is stored in such a tangible non-temporary information storage medium and provided, the program signal can be provided as a download service by using the electric telecommunication line such as the Internet.

[0144] Further, a part or all of at least one processing function of the linear structure extraction function or the learning function described in the embodiment is provided as an application server, and services that provide the processing function can be performed through the electric telecommunication line.

<<About Hardware Configuration of Each Processing Unit>>

[0145] The hardware structures of the processing units that execute various pieces of processing, such as the first neural network 41, the DCNN 411, the RPN 412, the second neural network 42, the R-CNN 423, which are in Fig. 4, the data acquisition unit 102, the learning model 52, the first error calculation unit 110, the second error calculation unit 112, and the optimizer 114, which are in Fig. 17, are various processors as shown below, for example.

[0146] The various processors include the CPU that is a general-purpose processor executing the program and functioning as the various processing units, the GPU that is a processor specialized in the image processing, a programmable logic device (PLD) that is a processor whose circuit configuration can be changed after manufacture, such as a field programmable gate array (FPGA), and a dedicated electric circuit that is a processor having a circuit configuration that is designed for exclusive use in order to execute specific processing, such as an application specific integrated circuit (ASIC).

[0147] One processing unit may be configured by one of these various processors or may be configured by two or more processors of the same type or different types. For example, one processing unit may be configured by a plurality of FPGAs, a combination of the CPU and the FPGA, or a combination of the CPU and the GPU. Alternatively, a plurality of processing units may be configured by one processor. As an example of the configuration of a plurality of processing units with one processor, first, as represented by a computer such as a client or a server, there is a form in which one processor is configured by a combination of one or more CPUs and the software and the processor functions as a plurality of processing units. Second, as represented by a system on chip (SoC), there is a form in which a processor is used in which the function of the entire system including a plurality of processing units is realized by a single integrated circuit (IC) chip. As described above, the various processing units are configured by one or more of the above various processors as a hardware structure.

[0148] Furthermore, the hardware structure of these various processors is, more specifically, an electric circuit (circuitry) in which the circuit elements such as semiconductor elements are combined.

«Effect of Embodiment»

[0149]

(1) According to the present embodiment, it is possible to extract the linear structure from the image.

(2) According to the present embodiment, the element points of the linear structure are directly detected, and thus it is easy to reconstruct the graph structure.

(3) According to the present embodiment, the correct to each training image is defined by the data indicating the positional information of the center line, it is easy to create the learning data.

<<Other Application Examples>>

[0150] The technique of the linear structure extraction processing according to the present disclosure is not limited to the CT image, and can be applied to various three-dimensional tomographic images. For example, the learning method may be applied to a magnetic resonance (MR) image acquired by a magnetic resonance imaging (MRI) device, a positron emission tomography (PET) image acquired by a PET device, an optical coherence tomography (OCT) image acquired by an OCT device, a three-dimensional ultrasound image acquired by a three-dimensional ultrasound imaging device, and the like.

[0151] Further, the technique of the linear structure extraction processing according to the present disclosure is limited

to the three-dimensional tomographic image, and can be applied to various two-dimensional images. For example, the image, which is the processing target, may be a two-dimensional X-ray image. Further, the technique of the linear structure extraction processing according to the present disclosure is not limited to the medical image, and can be applied to various images such as a normal camera image. For example, the technique of the present disclosure can be applied in a case in which the crack is detected from the image of the building or the like handled in Gahayun Suh, Young-Jin Cha "Deep faster R-CNN-based automated detection and localization of multiple types of damage" Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018.

«Modification Example»

[0152]

[1] The type of the anchor may be one type depending on the shape and/or the size of the target object to be detected.
[2] It is also possible to adopt a form in which the calculation of the deformation magnification of the anchor is not performed in the RPN 412, such as a case in which the size of the target object does not matter.

«Others»

[0153] The matters described in the configuration or the modification example described in the embodiment described above can be used in combination as appropriate, and a part of the matters can be replaced. It is needless to say that the present invention is not limited to the embodiment described above and various modifications can be made without departing from the spirit of the present invention.

Explanation of References

[0154]

40:	Faster R-CNN
41:	first neural network
42:	second neural network
50:	linear structure extraction device
52:	learning model
71:	first anchor
72:	second anchor
81, 82, 83:	anchor
100:	learning device
102:	data acquisition unit
110:	first error calculation unit
112:	second error calculation unit
114:	optimizer
150:	learning data storage unit
410:	first subnetwork
411:	DCNN
412:	RPN
414, 416, 417, 418:	convolutional layer
420:	second subnetwork
423:	R-CNN
424:	ROI pooling layer
426, 427:	fully bonded layer
428:	softmax layer
800:	computer
810:	storage
812:	communication unit
814:	input device
816:	display device
818:	bus
A1, A2, A3:	anchor
BV:	blood vessel

Car:	coronary artery route
CLbv:	center line
CP, CP34:	center coordinate
NP, NP34:	closest point
Nd:	node
Eg:	edge
HVR1:	image
HVR2:	image
IMipt:	input image
FM:	feature map
RP:	candidate region
LD(i):	learning data
PLQ:	plaque
px:	pixel
pxfm:	pixel
u: unit S50 to S58:	step of linear structure extraction processing
S61 to S67:	step of candidate region generation processing
S202 to S210:	step of learning processing

Claims

1. A linear structure extraction device that extracts element points which constitute a linear structure from an image, the device comprising:

a learning model that is learned to receive an input of the image and output, as a prediction result, one or more element points which constitute the linear structure from the image,
wherein the learning model includes

a first processing module that receives the image and generates a feature map representing a feature amount of the image by convolution processing, and
a second processing module that calculates a shift amount from a unit center point to the element point of the linear structure closest to the unit center point, for each unit obtained by dividing the feature map into a plurality of the units including regions having a predetermined size in a grid pattern.

2. The linear structure extraction device according to claim 1,
wherein the second processing module

disposes anchors, which are one or more reference shape regions having predetermined shape and size, in each of the units, and
performs, for each unit, the convolution processing by using the feature amount at a position of the unit to calculate the shift amount for moving an anchor center point of the anchor to a closest point, which is the element point of the linear structure closest to the anchor center point, and a score for determining whether or not the linear structure is present in the anchors.

3. The linear structure extraction device according to claim 2,
wherein the reference shape region is a rectangular region in a case in which the image is a two-dimensional image, and is a rectangular parallelepiped region in a case in which the image is a three-dimensional image.

4. The linear structure extraction device according to claim 2 or 3,

wherein the linear structure is a representative line of a region having a thickness in the image, and
a plurality of the anchors having different sizes are used in accordance with the thickness of the region having the thickness.

5. The linear structure extraction device according to any one of claims 2 to 4,

wherein the linear structure is a representative line of a region having a thickness in the image, and

the second processing module is learned to change the size of the anchor in accordance with the thickness of the region having the thickness, which is a target.

6. The linear structure extraction device according to any one of claims 2 to 5,

wherein the linear structure is a representative line of a region having a thickness in the image, and
the second processing module is learned to calculate, for each anchor, a deformation magnification of the anchor in a direction of at least one side of the anchor in accordance with the thickness of the region having the thickness around the closest point.

7. The linear structure extraction device according to any one of claims 4 to 6,

wherein the region having the thickness is a tubular structure, and
the representative line is a center line along a route of the tubular structure.

8. The linear structure extraction device according to any one of claims 1 to 7,

wherein each of the first processing module and the second processing module is configured by a neural network,
the first processing module is configured by a convolutional neural network including a plurality of convolutional layers, and
the second processing module includes a convolutional layer different from that of the first processing module and is configured by a region proposal network that predicts a candidate region including the linear structure from the feature map.

9. The linear structure extraction device according to any one of claims 1 to 8, further comprising:
a third processing module that is learned to perform classification of each point with respect to the element points of the linear structure predicted by the second processing module.

10. The linear structure extraction device according to claim 9,
wherein a class classified by the third processing module includes at least one of a root, a branch, a tip, or a point on a twig in a tree structure of graph theory.

11. The linear structure extraction device according to claim 9,

wherein the linear structure is a center line along a route of a blood vessel, and
a class classified by the third processing module includes a specific anatomical name in a vascular structure.

12. The linear structure extraction device according to claim 9,

wherein the linear structure is a center line along a route of a trachea, and
a class classified by the third processing module includes a specific anatomical name in a tracheal structure.

13. The linear structure extraction device according to any one of claims 9 to 12,

wherein the third processing module is configured by a neural network, and
the third processing module includes

a region-of-interest pooling layer that cuts out a local image of the anchor including the element point predicted by the second processing module from the feature map and deforms the local image to a fixed size, and
at least one of a convolutional layer or a fully bonded layer to which the local image deformed into the fixed size is input.

14. A linear structure extraction method that extracts element points which constitute a linear structure from an image, the method comprising:

by using a learning model that is learned to receive an input of the image and output, as a prediction result, one or more element points which constitute the linear structure from the image,

receiving the input of the image to the learning model;
generating a feature map representing a feature amount of the image by performing convolution processing on the input image by using a first processing module; and
dividing the feature map into a plurality of units including regions having a predetermined size in a grid pattern and calculating, for each unit, a shift amount from a unit center point to the element point of the linear structure closest to the unit center point by using a second processing module.

15. The linear structure extraction method according to claim 14, further comprising:
deleting, from a point cloud of a plurality of the element points predicted by the plurality of units, a part of excess element points close to each other by less than a first interval, which is half of the size of the unit as a standard, and selecting and leaving the element points at a degree of the first interval.

16. The linear structure extraction method according to claim 14,

wherein the linear structure is a representative line of a region having a thickness in the image, and
the linear structure extraction method further comprises deleting, from a point cloud of a plurality of the element points predicted by the plurality of units, a part of excess element points close to each other by less than a second interval, which is half of the thickness as a standard, and selecting and leaving the element points at a degree of the second interval.

17. The linear structure extraction method according to any one of claims 14 to 16, further comprising:
deleting, from a point cloud of a plurality of the element points predicted by the plurality of units, an isolation point around which other points are not present within a distance of a predetermined threshold value.

18. A program that causes a computer to realize a function of extracting element points which constitute a linear structure from an image, the program causing the computer to realize:

a function of receiving an input of the image;
a function of performing convolution processing on the input image by using a first processing module and generating a feature map representing a feature amount of the image; and
a function of dividing the feature map into a plurality of units including regions having a predetermined size in a grid pattern and predicting, for each unit, a shift amount from a center point of the unit to a closest element point of the linear structure by using a second processing module.

19. A non-temporary computer-readable recording medium that causes a computer to execute the program according to claim 18 in a case in which a command stored in the recording medium is read by the computer.

20. A learned model that is learned to output, as a prediction result, one or more element points which constitute a linear structure from an input image, the model comprising:

a first processing module that receives the image and generates a feature map representing a feature amount of the image by convolution processing; and
a second processing module that calculates a shift amount from a unit center point to the element point of the linear structure closest to the unit center point, for each unit obtained by dividing the feature map into a plurality of the units including regions having a predetermined size in a grid pattern.

21. The learned model according to claim 20,
wherein parameters of networks which configure the first processing module and the second processing module are decided by executing machine learning by using a plurality of pieces of learning data in which a training image and positional information of the linear structure included in the training image are combined.

22. The learned model according to claim 21,

wherein the linear structure is a representative line of a region having a thickness in the image, and
the learning data further includes thickness information of the region having the thickness included in the training image.

FIG. 1

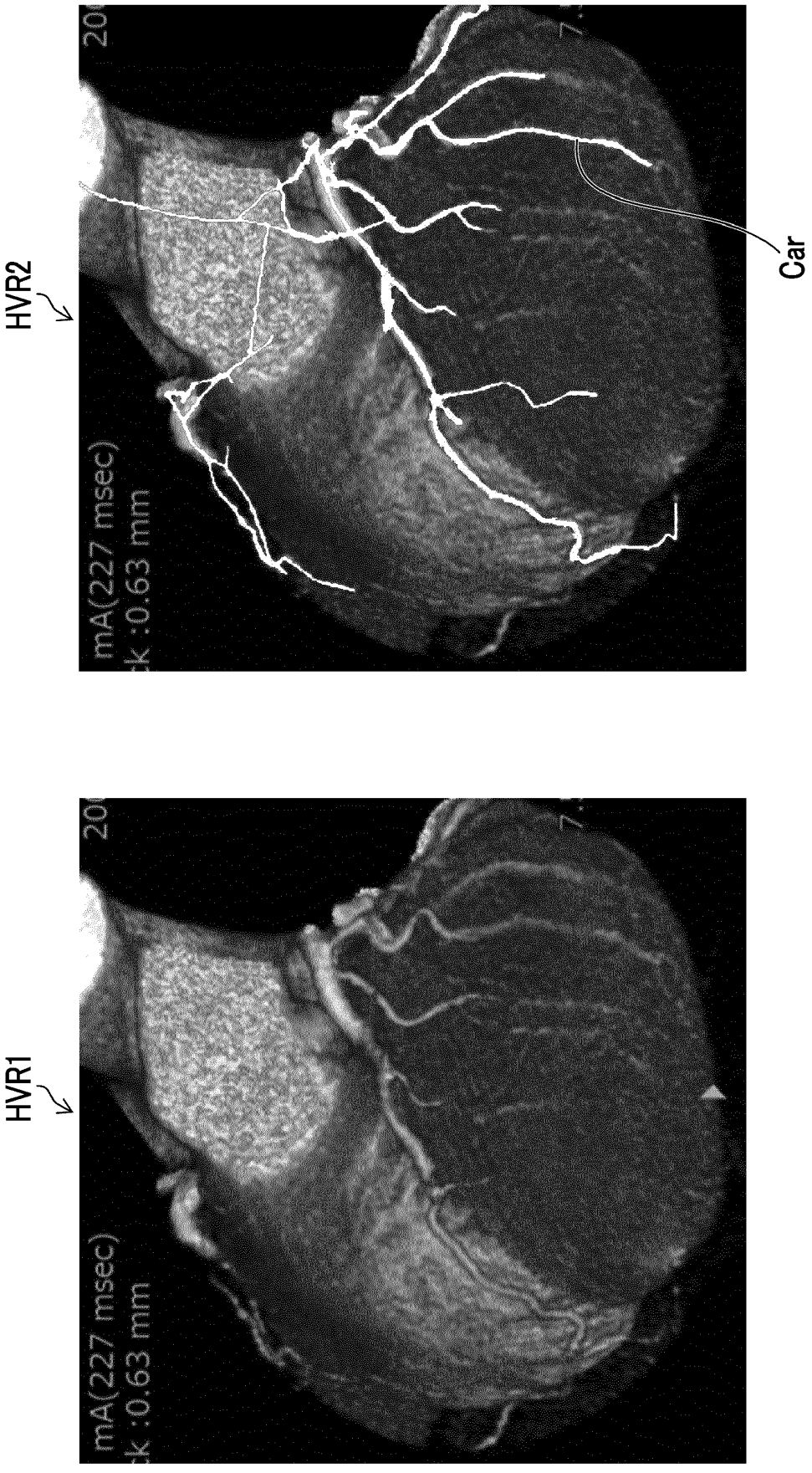


FIG. 2

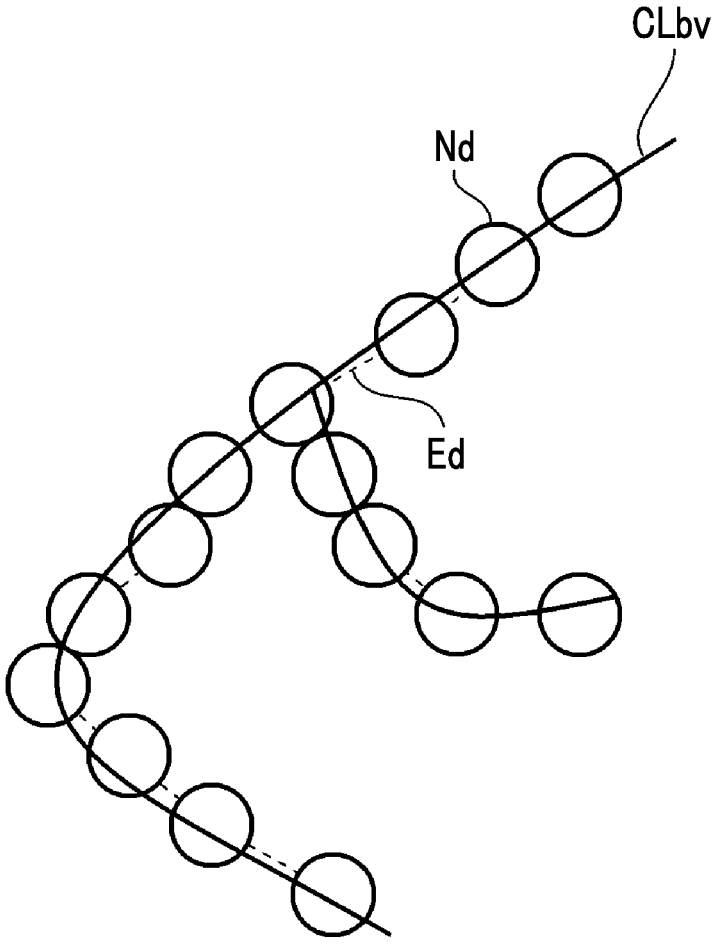


FIG. 3

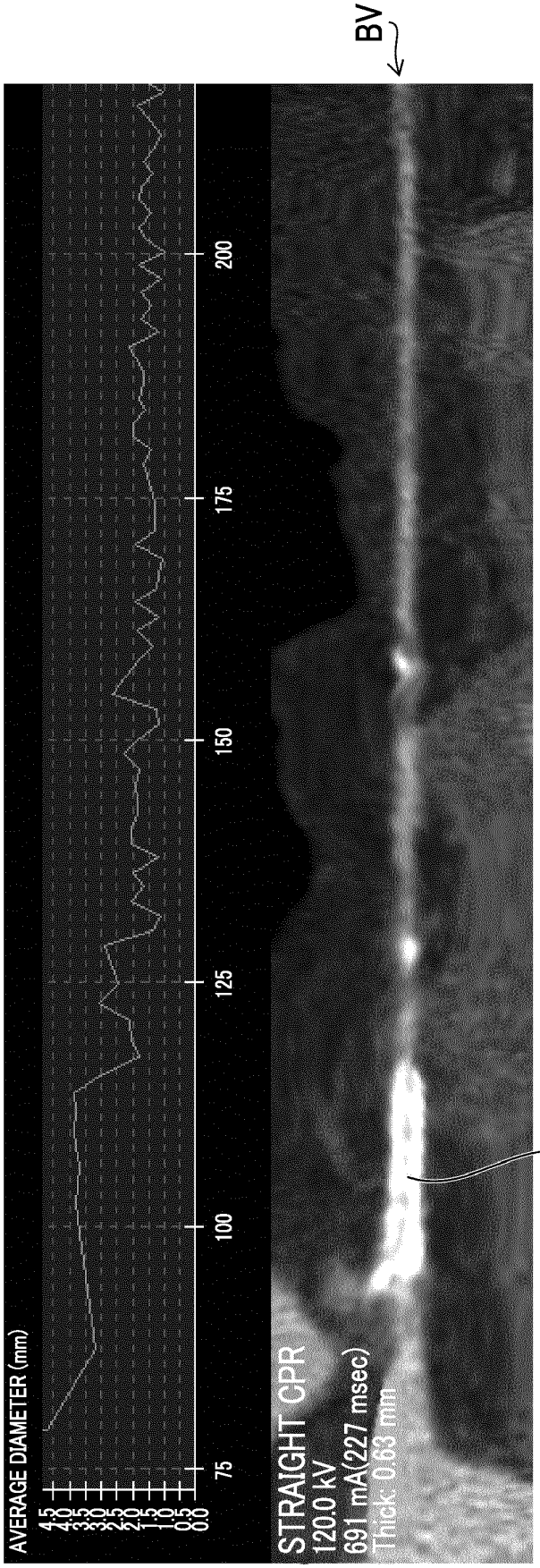

BV
PLQ
AVERAGE DIAMETER (mm)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
75
100
125
150
175
200
STRAIGHT CPR
120.0 kV
691 mA(227 msec)
Thick 0.63 mm

FIG. 4

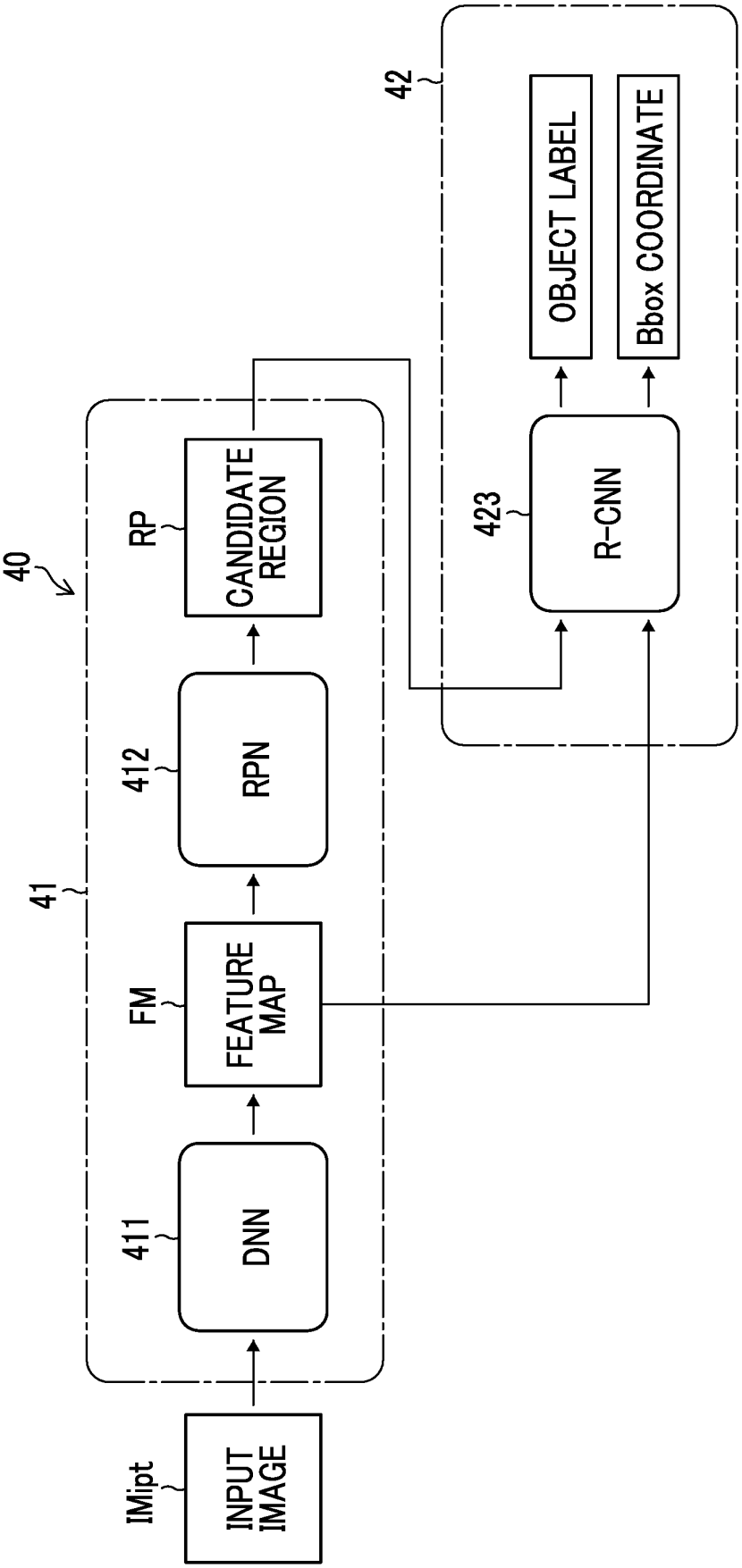

IMipt
INPUT IMAGE
411
DNN
FM
FEATURE MAP
412
RPN
RP
CANDIDATE REGION
40
41
42
423
R-CNN
OBJECT LABEL
Bbox COORDINATE

FIG. 5

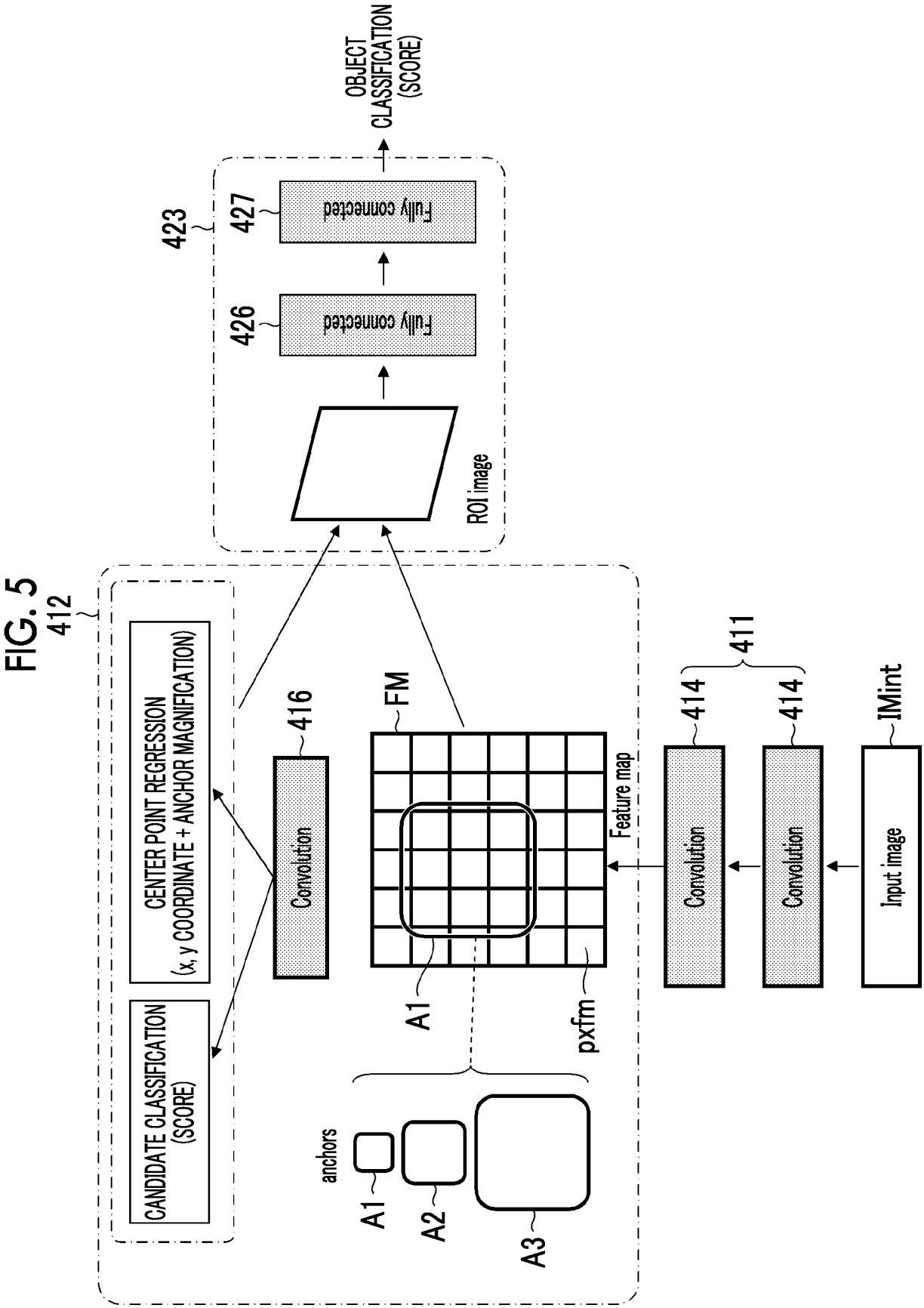

OBJECT CLASSIFICATION (SCORE)
423
427
Fully connected
426
Fully connected
ROI image
412
CENTER POINT REGRESSION (x, y COORDINATE + ANCHOR MAGNIFICATION)
CANDIDATE CLASSIFICATION (SCORE)
416
Convolution
FM
Feature map
411
414
414
Convolution
Convolution
IMint
Input image
A1
pxfm
anchors
A1
A2
A3

FIG. 6

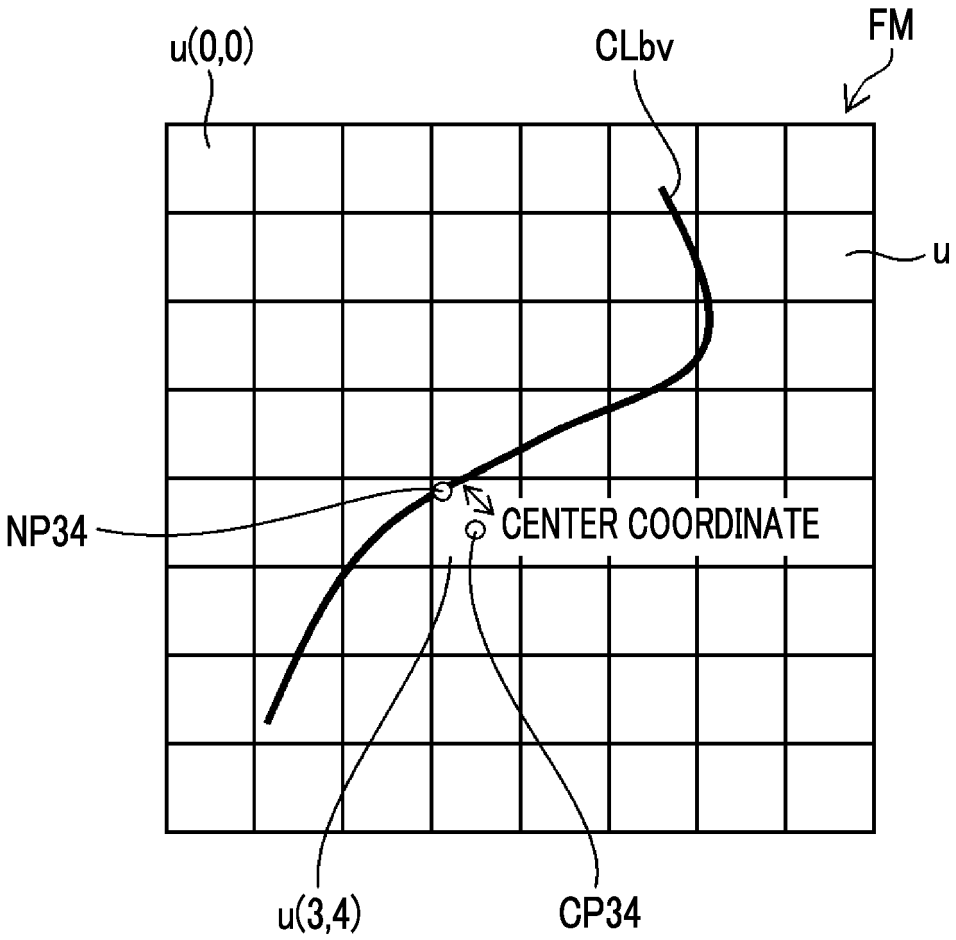


FIG. 7

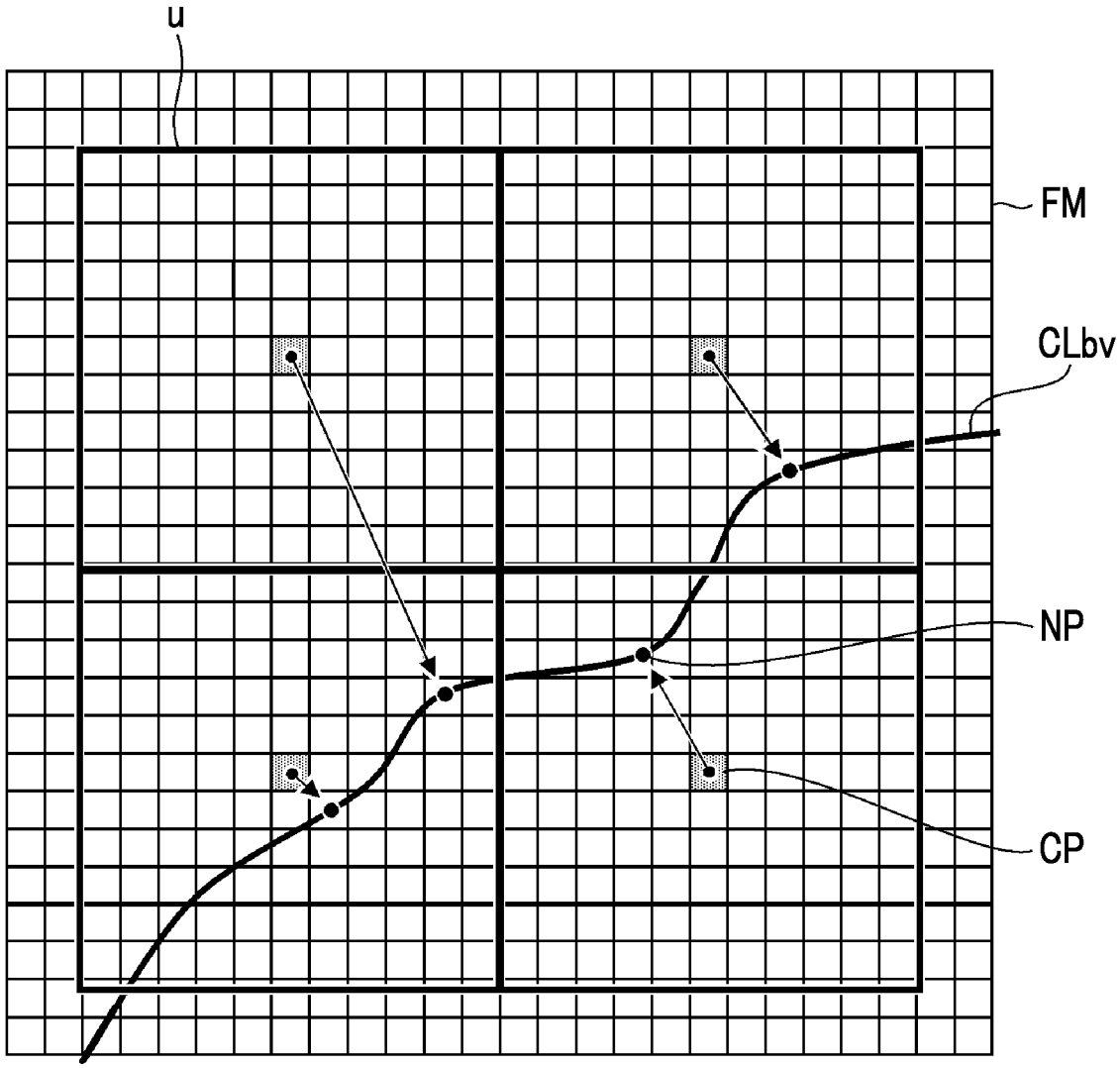


FIG. 8

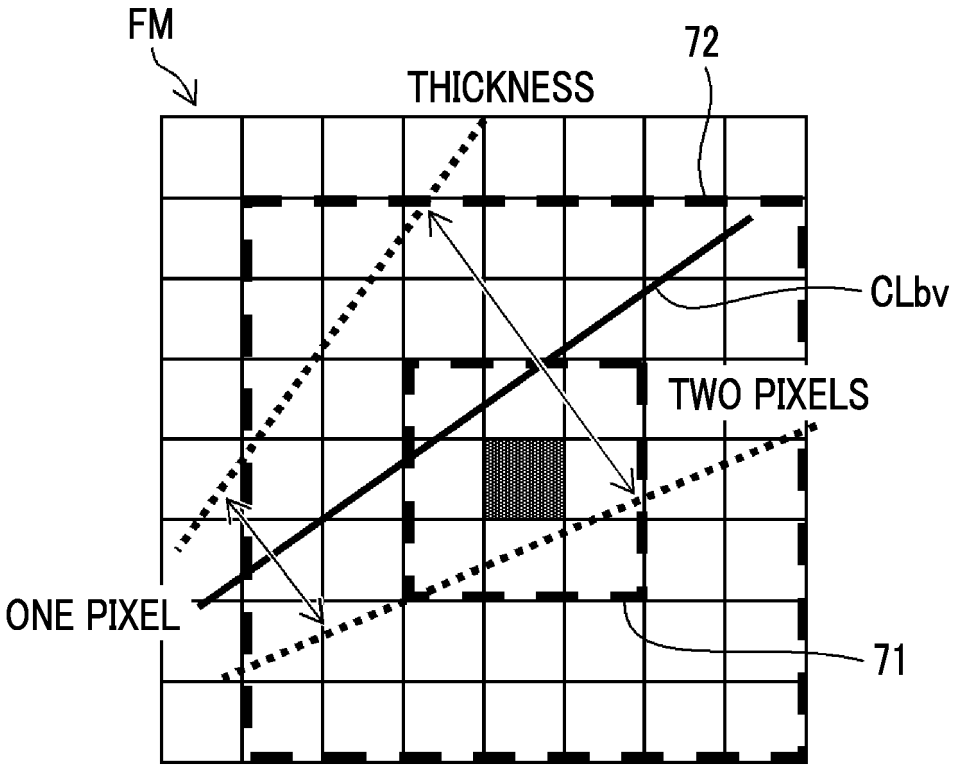


FIG. 9

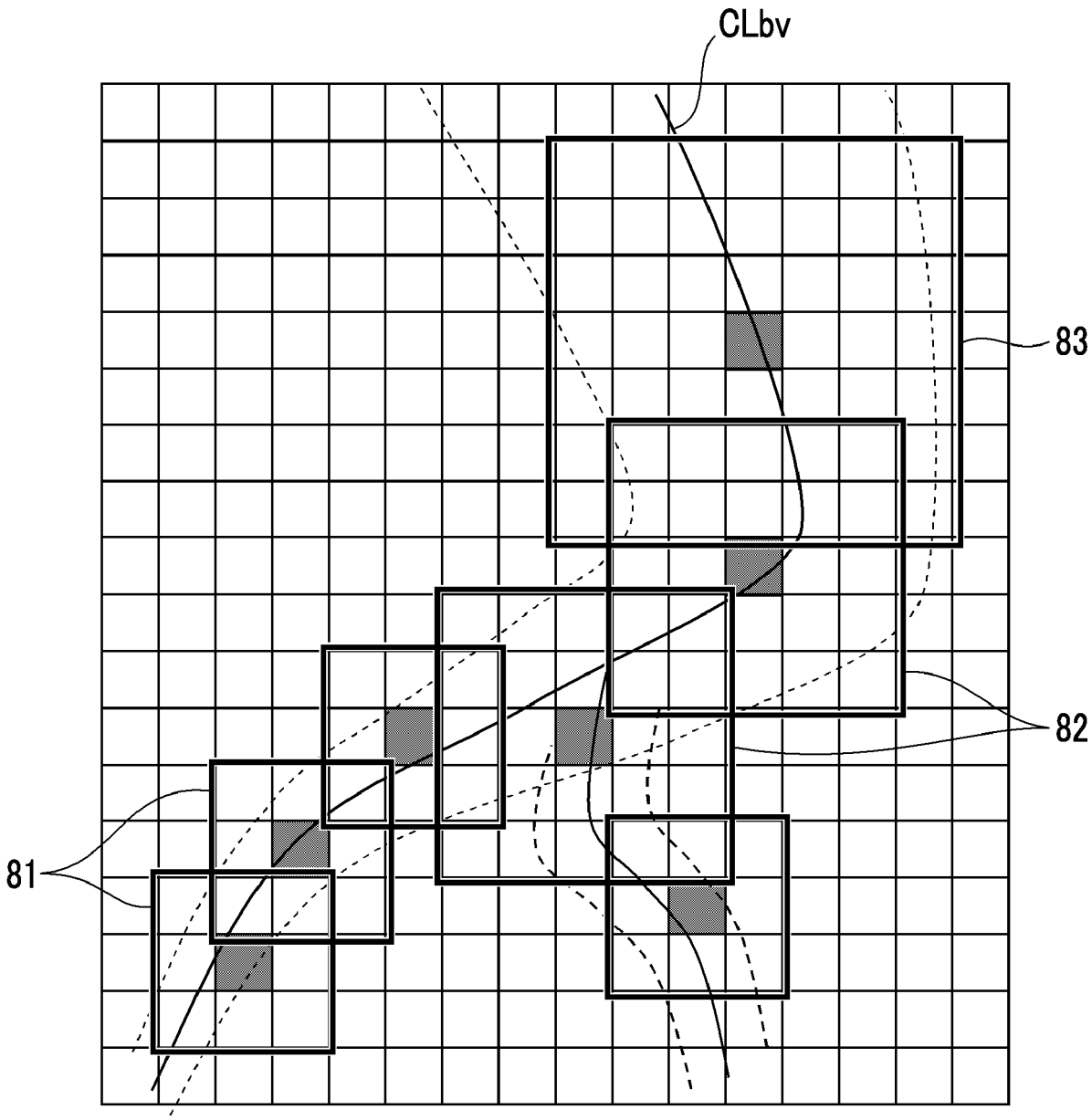


FIG. 10

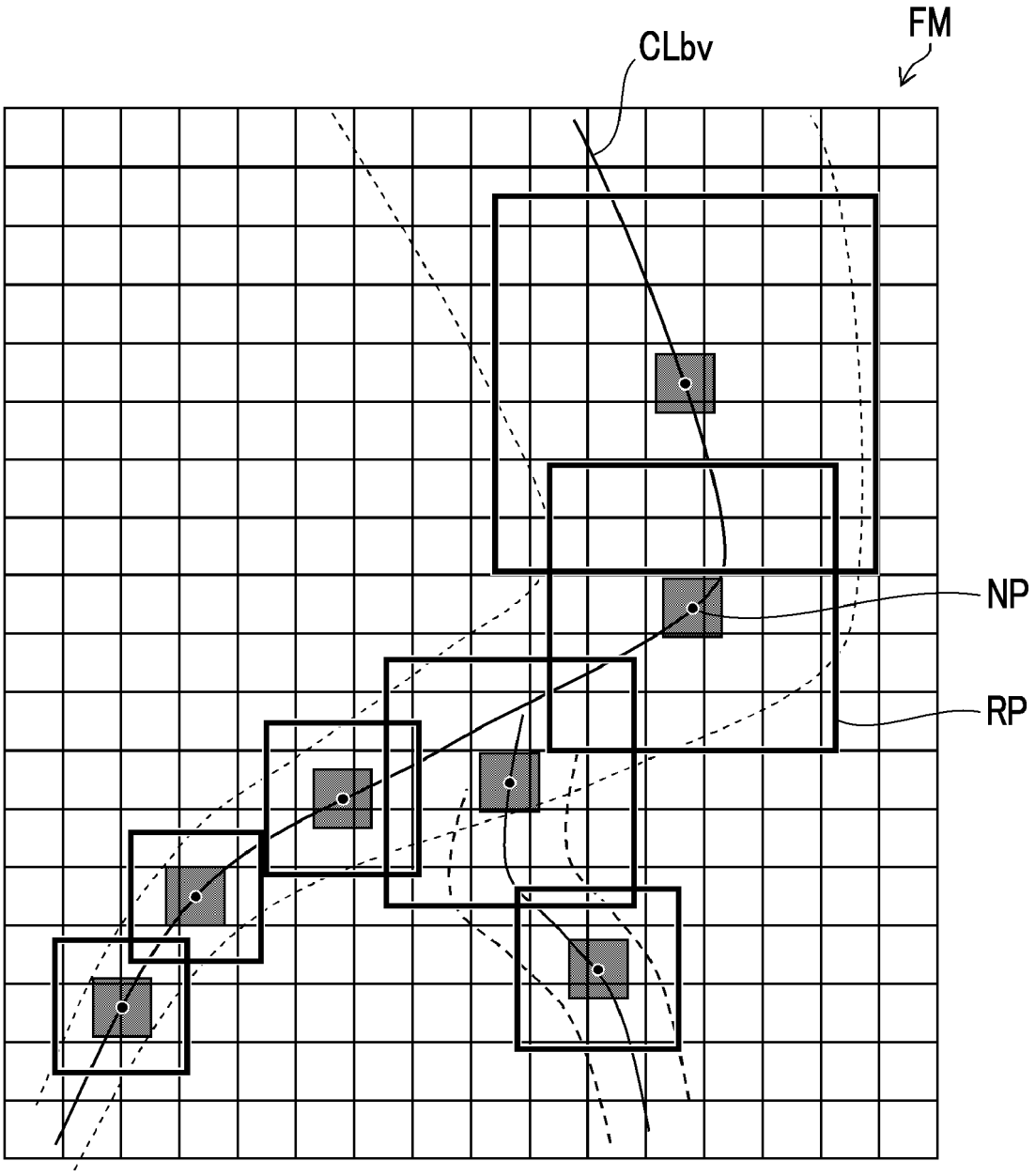


FIG. 11

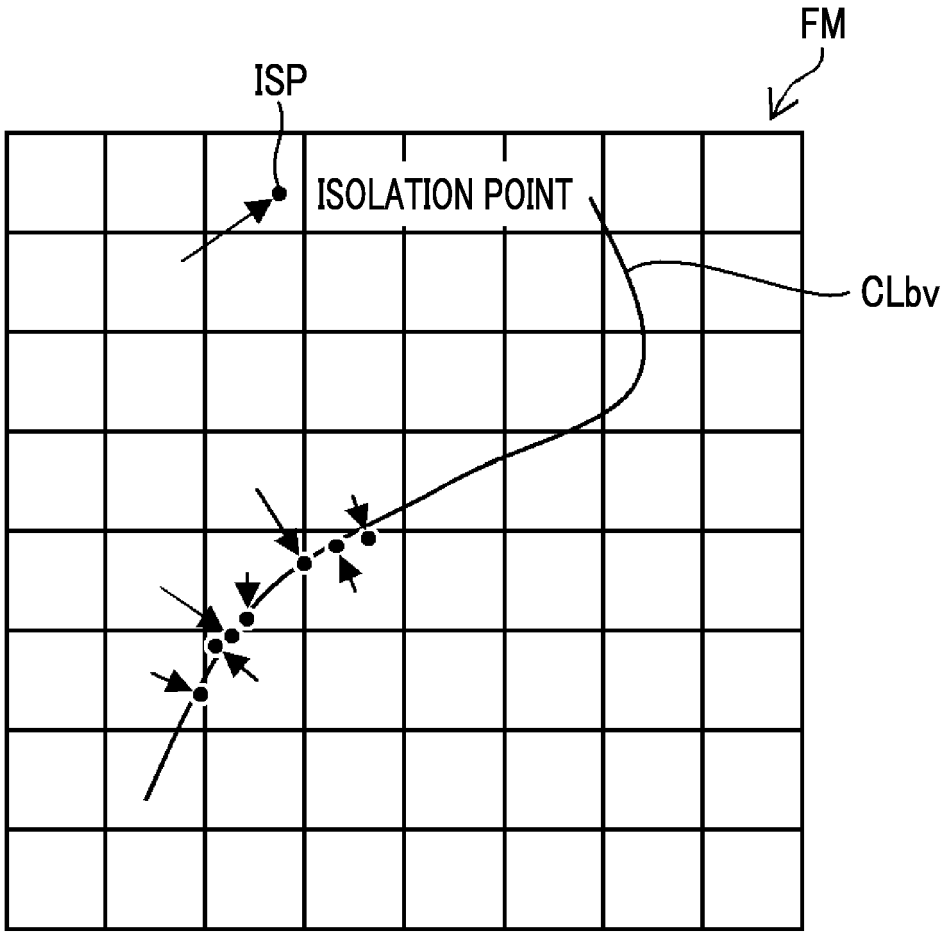


FIG. 12

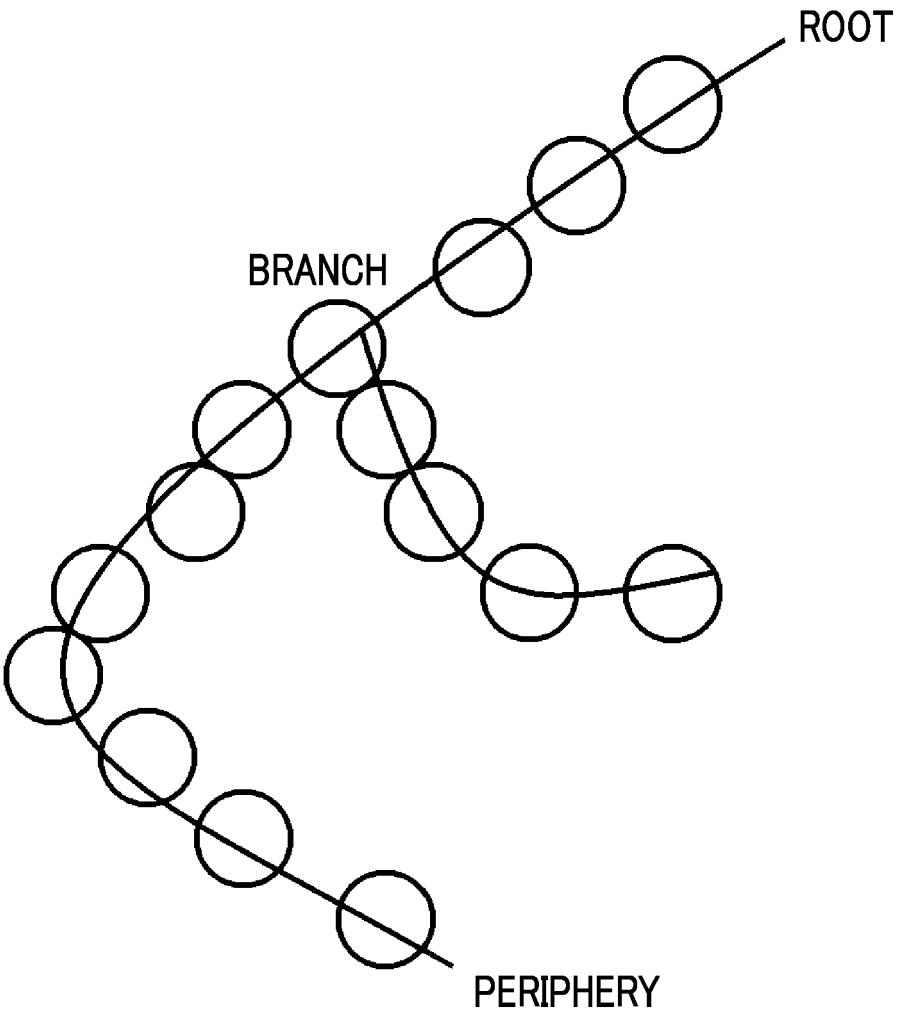


FIG. 13

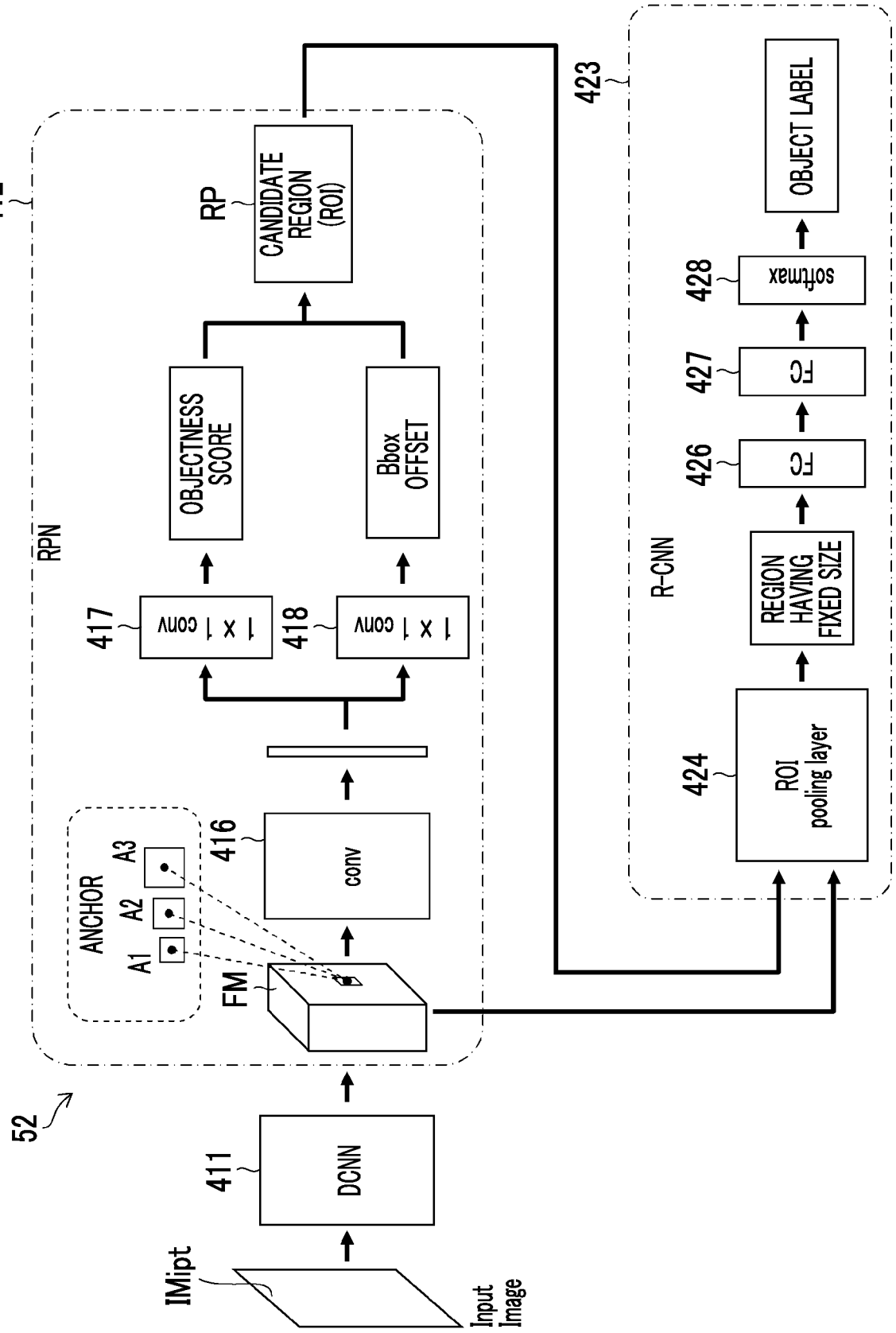

IMipt
Input Image
411
DCNN
52
412
RPN
ANCHOR
A1
A2
A3
FM
416
conv
417
1 x 1 conv
418
1 x 1 conv
OBJECTNESS SCORE
Bbox OFFSET
RP
CANDIDATE REGION (ROI)
423
R-CNN
424
ROI pooling layer
REGION HAVING FIXED SIZE
426
FC
427
FC
428
softmax
OBJECT LABEL

FIG. 14

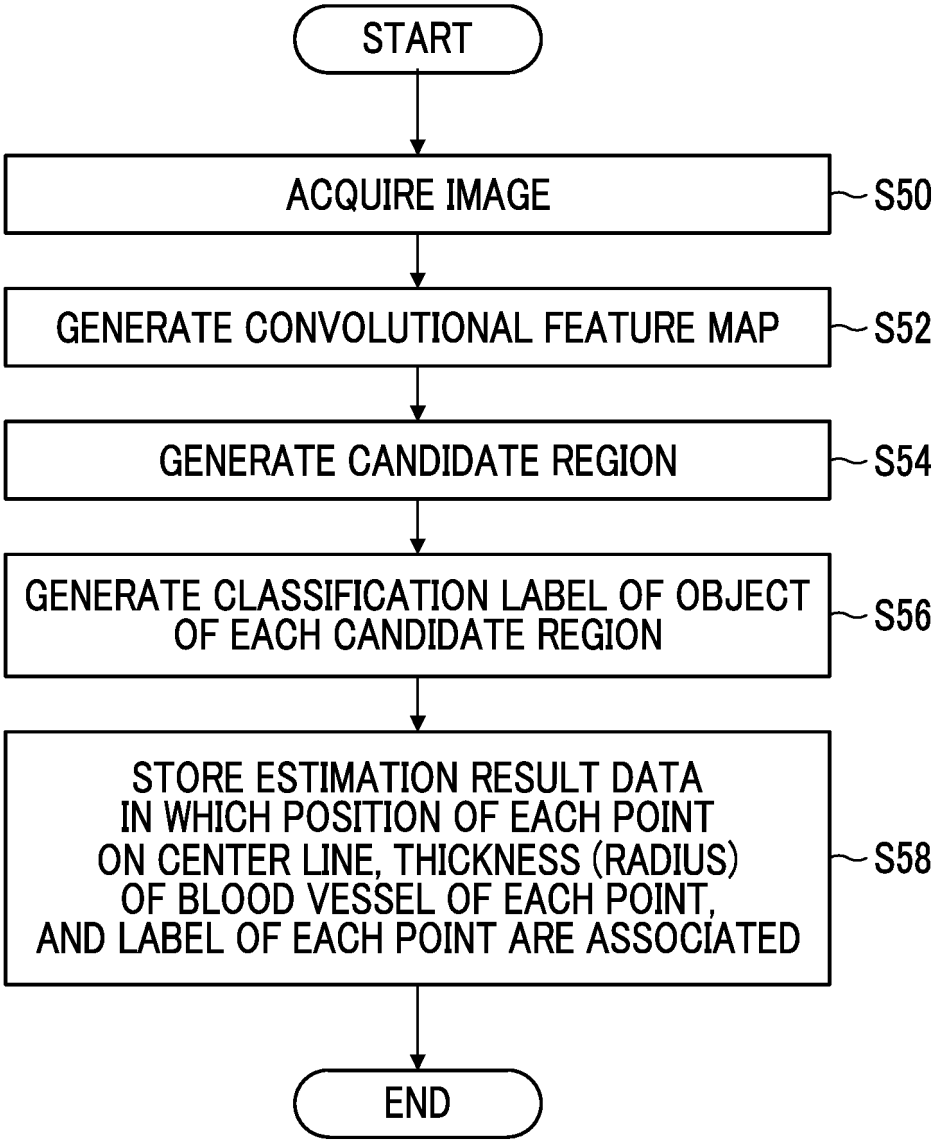

START
ACQUIRE IMAGE
S50
GENERATE CONVOLUTIONAL FEATURE MAP
S52
GENERATE CANDIDATE REGION
S54
GENERATE CLASSIFICATION LABEL OF OBJECT OF EACH CANDIDATE REGION
S56
STORE ESTIMATION RESULT DATA IN WHICH POSITION OF EACH POINT ON CENTER LINE, THICKNESS (RADIUS) OF BLOOD VESSEL OF EACH POINT, AND LABEL OF EACH POINT ARE ASSOCIATED
S58
END

FIG. 15

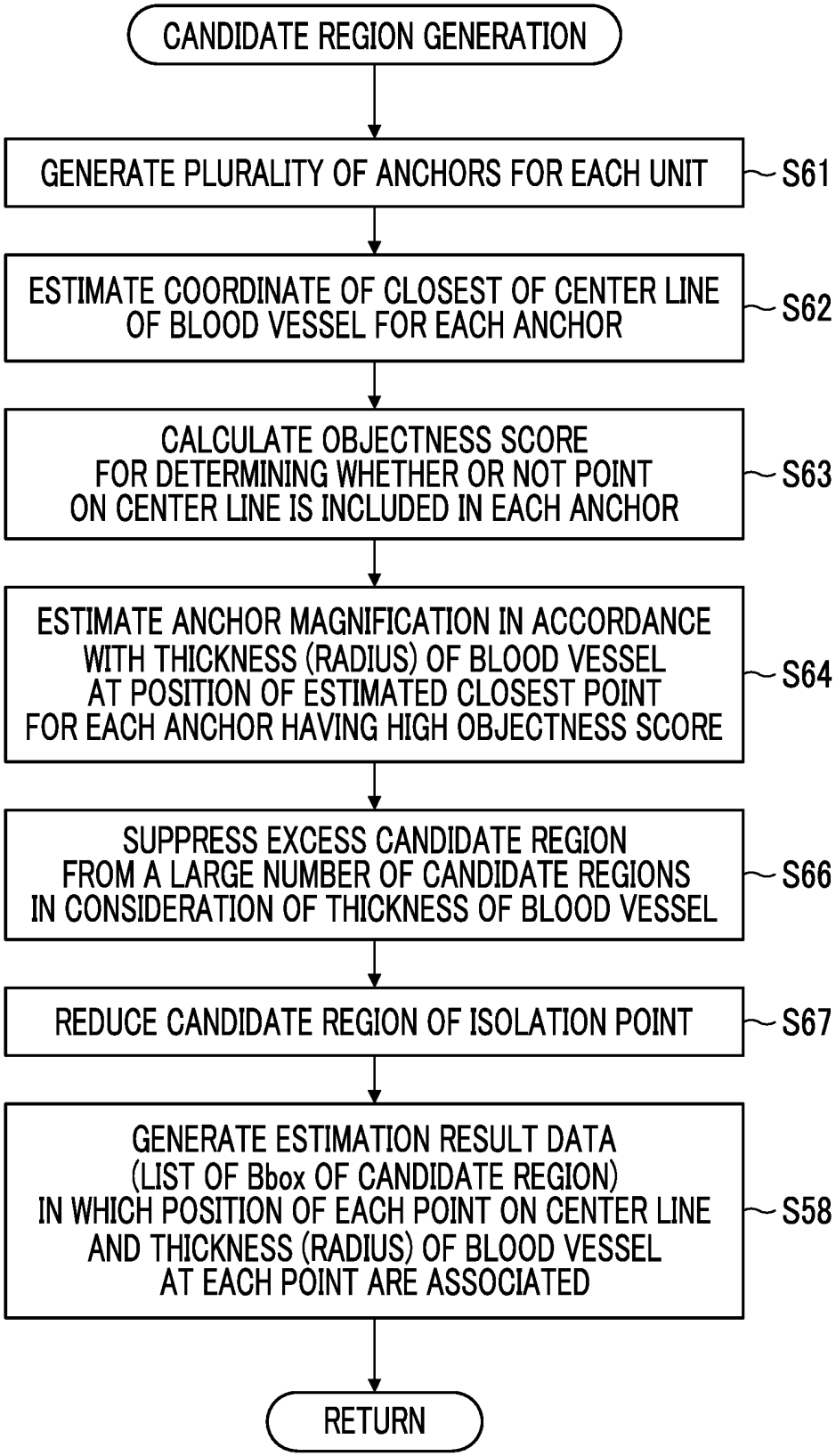

CANDIDATE REGION GENERATION
GENERATE PLURALITY OF ANCHORS FOR EACH UNIT
S61
ESTIMATE COORDINATE OF CLOSEST OF CENTER LINE OF BLOOD VESSEL FOR EACH ANCHOR
S62
CALCULATE OBJECTNESS SCORE FOR DETERMINING WHETHER OR NOT POINT ON CENTER LINE IS INCLUDED IN EACH ANCHOR
S63
ESTIMATE ANCHOR MAGNIFICATION IN ACCORDANCE WITH THICKNESS (RADIUS) OF BLOOD VESSEL AT POSITION OF ESTIMATED CLOSEST POINT FOR EACH ANCHOR HAVING HIGH OBJECTNESS SCORE
S64
SUPPRESS EXCESS CANDIDATE REGION FROM A LARGE NUMBER OF CANDIDATE REGIONS IN CONSIDERATION OF THICKNESS OF BLOOD VESSEL
S66
REDUCE CANDIDATE REGION OF ISOLATION POINT
S67
GENERATE ESTIMATION RESULT DATA (LIST OF Bbox OF CANDIDATE REGION) IN WHICH POSITION OF EACH POINT ON CENTER LINE AND THICKNESS (RADIUS) OF BLOOD VESSEL AT EACH POINT ARE ASSOCIATED
S58
RETURN

FIG. 16

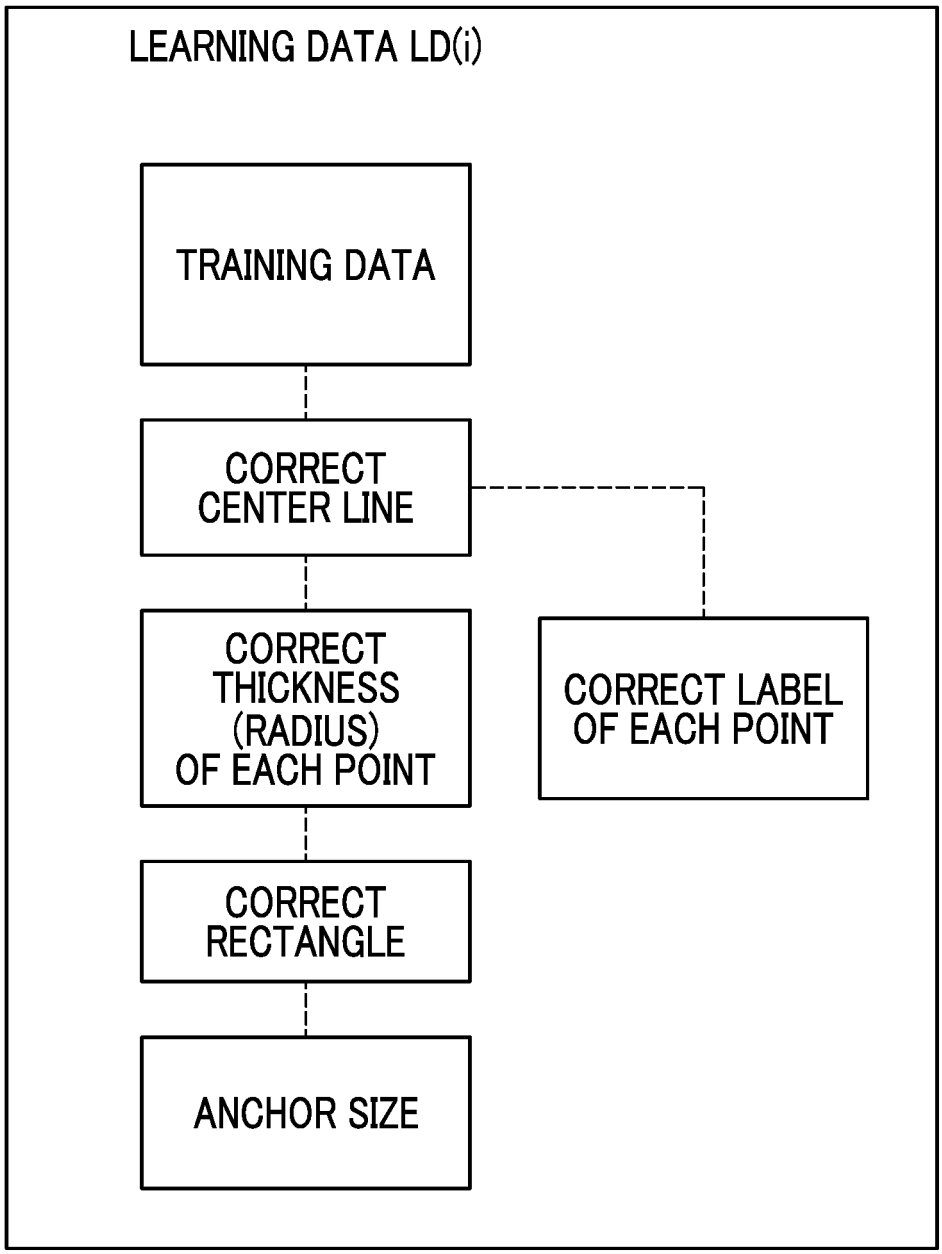

LEARNING DATA LD(i)
TRAINING DATA
CORRECT CENTER LINE
CORRECT THICKNESS (RADIUS) OF EACH POINT
CORRECT LABEL OF EACH POINT
CORRECT RECTANGLE
ANCHOR SIZE

FIG. 17

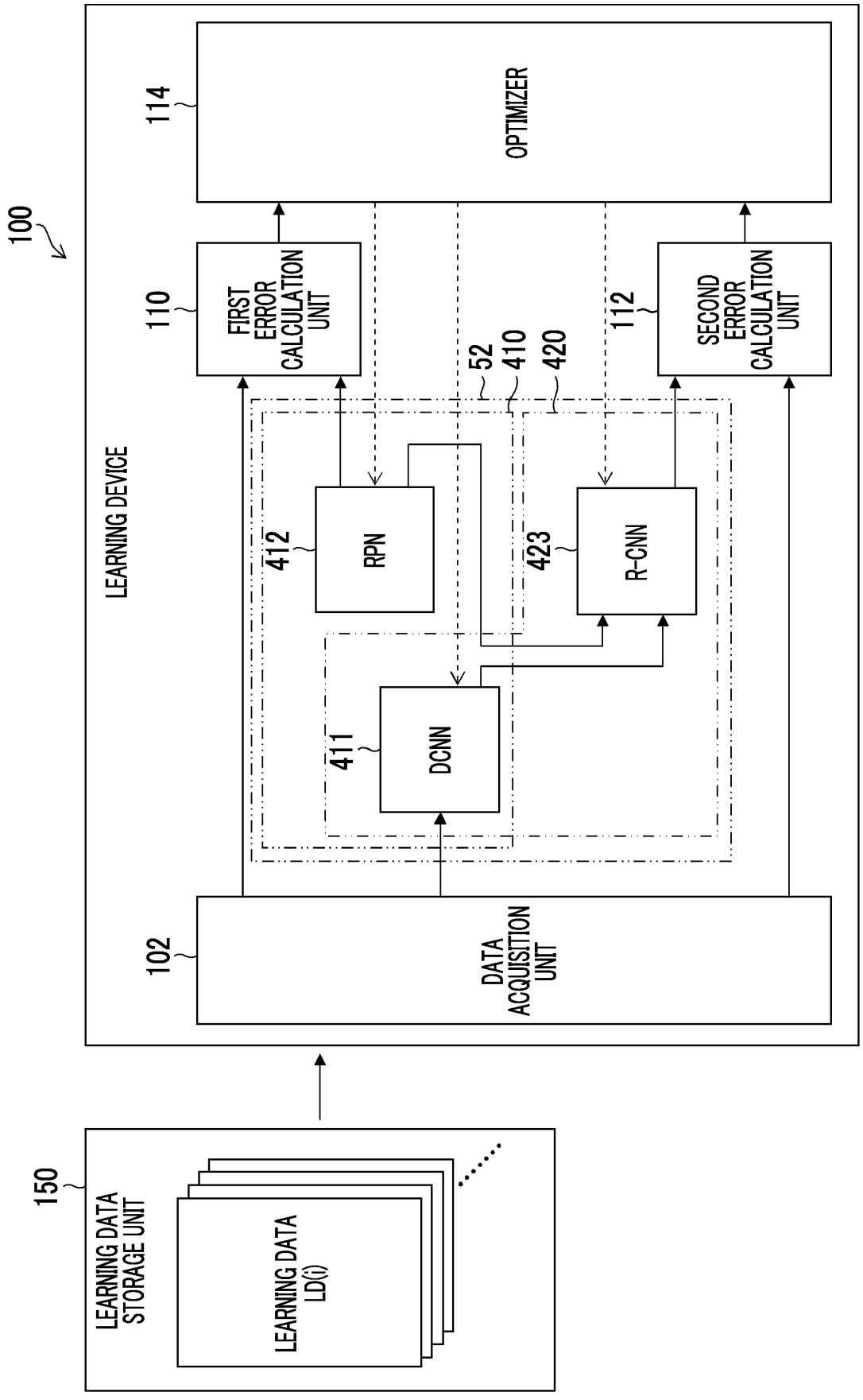

100
LEARNING DEVICE
114
OPTIMIZER
110
FIRST ERROR CALCULATION UNIT
112
SECOND ERROR CALCULATION UNIT
52
410
420
412
RPN
423
R-CNN
411
DCNN
102
DATA ACQUISITION UNIT
150
LEARNING DATA STORAGE UNIT
LEARNING DATA LD(i)

FIG. 18

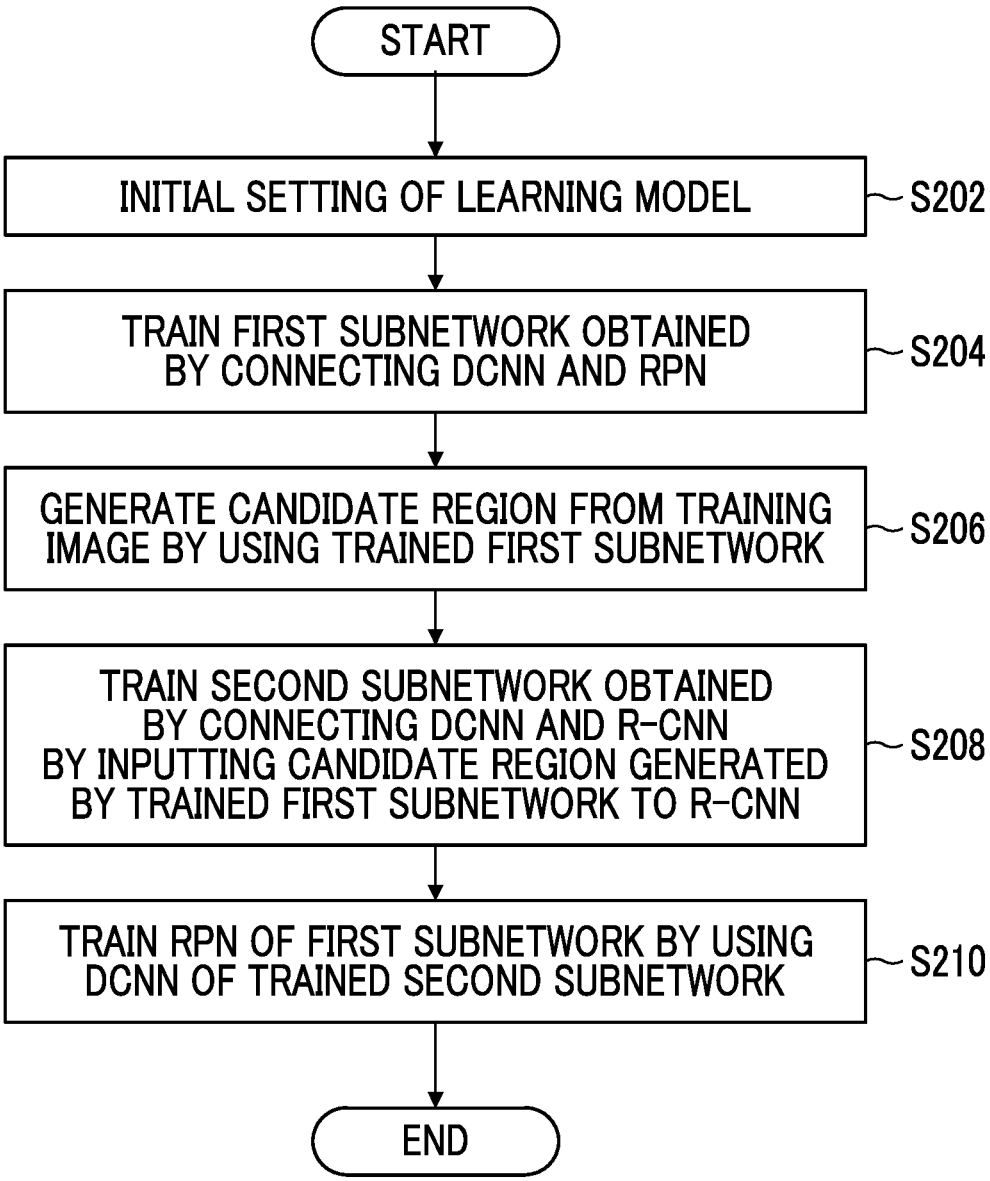
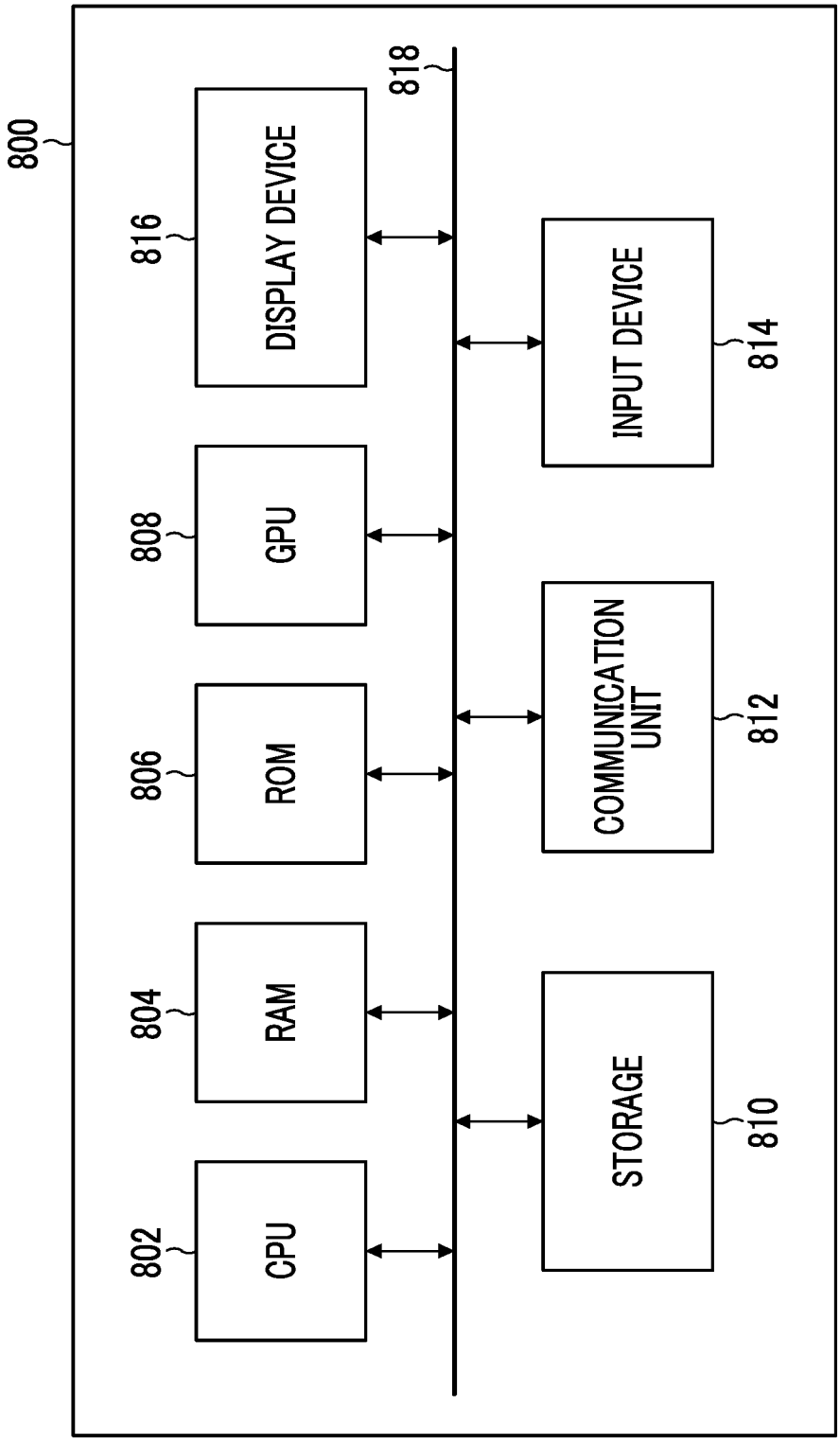

START
INITIAL SETTING OF LEARNING MODEL
S202
TRAIN FIRST SUBNETWORK OBTAINED BY CONNECTING DCNN AND RPN
S204
GENERATE CANDIDATE REGION FROM TRAINING IMAGE BY USING TRAINED FIRST SUBNETWORK
S206
TRAIN SECOND SUBNETWORK OBTAINED BY CONNECTING DCNN AND R-CNN BY INPUTTING CANDIDATE REGION GENERATED BY TRAINED FIRST SUBNETWORK TO R-CNN
S208
TRAIN RPN OF FIRST SUBNETWORK BY USING DCNN OF TRAINED SECOND SUBNETWORK
S210
END

FIG. 19


800
818
816
DISPLAY DEVICE
808
GPU
806
ROM
804
RAM
802
CPU
INPUT DEVICE
814
COMMUNICATION UNIT
812
STORAGE
810

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2020/013334

A. CLASSIFICATION OF SUBJECT MATTER
G01T 1/161(2006,01)i; G06T 7/00(2017,01)i; A61B 6/03(2006,01)i; A61B 5/055(2006,01)i
FI: G06T7/00 350C; G06T7/00 612; A61B6/03 360J; G01T1/161 A; A61B5/055 380; A61B6/03 360G
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G01T1/161; G06T7/00; A61B6/03; A61B5/055

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2020
Registered utility model specifications of Japan	1996-2020
Published registered utility model applications of Japan	1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2019-500110 A (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 10.01.2019 (2019-01-10) entire text, all drawings	1-22
A	JP 2018-534694 A (NEC LABORATORIES AMERICA, INC.) 22,11.2018 (2018-11-22) entire text, all drawings	1-22
A	US 10198671 B1 (SNAP INC.) 05.02.2019 (2019-02-05) entire text, all drawings	1-22

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 June 2020 (10.06.2020)	23 June 2020 (23.06.2020)

Name and mailing address of the ISA/	Authorized officer
Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Telephone No.

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application no.
PCT/JP2020/013334

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2019-500110 A	10 Jan. 2019	WO 2017/106645 A1 entire text, all drawings	
JP 2018-534694 A	22 Nov. 2018	US 2017/0124415 A1 entire text, all drawings	
US 10198671 B1	05 Feb. 2019	(Family: none)	

Form PCT/ISA/210 (patent family annex) (January 2015)

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 9858496 B **[0002] [0003] [0045] [0072] [0073]**

Non-patent literature cited in the description

- **REN, SHAOQING et al.** Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems,* 2015 **[0002] [0003] [0045] [0065] [0072] [0073]**
- **GAHAYUN SUH ; YOUNG-JIN CHA.** Deep faster R-CNN-based automated detection and localization of multiple types of damage. *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems,* 2018 **[0002] [0004] [0151]**